AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 112:103–128 (2000)

Craniofacial Variation in *Homo habilis:* An Analysis of the Evidence for Multiple Species

JOSEPH M.A. MILLER
Department of Pathology and Laboratory Medicine, University of California, Los Angeles, California 90095-1732

KEY WORDS paleoanthropology; hominids; early Homo

ABSTRACT The question of heterogeneity in the *Homo habilis* sample continues to be controversial. Various lines of evidence have been used to reject the null hypothesis of intraspecific variation. This evidence derives from analyses of endocranial volume variation, probability estimates of sexual dimorphism, facial variation, cranial angles, CV analysis of craniofacial variation, the multivariate pattern of sexual dimorphism, the pattern of variability (CV) profiles, distance data using exact randomization methods, and various kinds of quantitative ordinations of fossils. Although consensus is lacking as to how the *H. habilis* sample is to be split, there is a growing perception that the degree of variation among the fossils is too great and the pattern of variation is too different to be explained by intraspecific variation. This has resulted in the recognition of new species such as "*Homo rudolfensis.*" The present study critically examines the evidence commonly cited as the basis for recognizing multiple species in the extended *H. habilis* hypodigm. Reanalysis and reinterpretation of these data indicates that: (1) the degree of variation in the *H. habilis* sample is typical of modern hominoids, and (2) the pattern of variation among specimens of the *H. habilis* sample is consistent with intraspecific variation. Thus, at present, there is no sound basis to reject the null hypothesis of intraspecific variation as an adequate explanation of the morphological variation seen among specimens of the extended *H. habilis* sample. If multiple species are indeed represented, then their presence has not yet been satisfactorily demonstrated. Am J Phys Anthropol 112:103–128, 2000.

Considerable debate surrounds the interpretation of the morphological variation observed among fossil specimens attributed to *Homo habilis.* Some view this variation as indicative of a sexually dimorphic, polytypic species (Howell, 1978; White et al., 1983; Johanson et al., 1987; Tobias, 1987, 1991). Others view it as evidence of multiple species (Walker and Leakey, 1978; Wood, 1978, 1985, 1991, 1992a,b, 1993; Stringer, 1986; Chamberlain, 1987; Lieberman et al., 1988; Groves, 1989; Rightmire, 1993; Kramer et al., 1995; Grine et al., 1996; Donnelly, 1996).

Though there is little agreement on how the extended *H. habilis* sample should be split, there seems to be an "increasing consensus" (cf. Lieberman et al., 1996) that the null hypothesis of intraspecific variation should be rejected. Some have claimed that the degree of variation in the *H. habilis* sample is too great for a single species (Wood, 1985; Stringer, 1986; Lieberman et al., 1988; Donnelly, 1996; Grine et al., 1996).

Grant sponsors: National Science Foundation, Smithsonian Predoctoral Fellowships.

*Correspondence to: Joseph M.A. Miller, Ph.D., Department of Pathology and Laboratory Medicine, University of California, Los Angeles, CHS 52-068 mc: 173216, 10833 Le Conte Ave, Los Angeles, CA 90095-1732. E-mail: jmmiller@mednet.ucla.edu

Received 30 April 1997; accepted 10 October, 1998.

Others have claimed that the pattern of variation is not consistent with intraspecific variation (Stringer, 1986; Wood, 1991, 1993; Kramer et al., 1995).

This study examines the evidence that has formed the basis for rejecting the single species (null) hypothesis for *H. habilis*. This evidence comes from studies of endocranial volume (ECV) variation (Wood, 1985; Stringer, 1986; but see Miller, 1991), probability estimates of sexual dimorphism (Lieberman et al., 1988), multivariate analysis of the facial region (Wood, 1985, citing Bilsborough and Wood, 1988), cranial angle variation (Stringer, 1986), coefficient of variance (CV) analysis of craniofacial measurements (Wood, 1991, 1993; Kramer et al., 1995), the multivariate pattern of sexual dimorphism (Wood, 1991, 1993), variability (CV) profiles (Kramer et al., 1995), distance measures using exact randomization procedures (Kramer et al., 1995; Grine et al., 1996; Donnelly, 1996), and quantitative ordinations of fossils (e.g., Rightmire, 1993; Grine et al., 1996).

The present study is limited to the more complete cranial specimens of the *H. habilis* sample that are at the center of the current debate. Mandibular, dental (isolated teeth), and postcranial data demand separate treatment (e.g., see Suwa et al., 1996 for a treatment of the early *Homo* sample of mandibular postcanine dentition). Also not addressed are qualitative character differences (e.g., Kimbel and Rak, 1993; Rightmire, 1993) since these proposed "taxonomic" differences have not yet been demonstrated to be anything other than individual variation.

The cranial specimens in this study are referred to as the *H. habilis* sample. This sample includes OH 7, OH 13, OH 16, OH 24, KNM-ER 1470, KNM-ER 1813, KNM-ER 1590, KNM-ER 3732, KNM-ER 1805, Stw 53, and SK 847. This sample results from the fact that these specimens were included in the taxon by one or more of the studies that are reconsidered here. It should not be taken as my endorsement of a hypodigm for *H. habilis*.

This study builds on my previous work on brain size variation in *H. habilis* (Miller, 1991). It also reports some preliminary results of on-going investigations (e.g., see Miller, 1990, 1994, 1995; Albrecht and Miller, 1997; Miller et al., 1997, 1998).

GENERAL APPROACH

The problem of species recognition in paleoanthropology continues to be the subject of much discussion (e.g., see Kimbel and Martin, 1993), and justifiably so, since studies of hominid phylogeny and other aspects of biology necessarily rest on a taxonomic foundation. For *H. habilis* and other disputed taxa, the problem is to correctly interpret the observed morphological variation in the fossil sample (Albrecht and Miller, 1993). This is done by determining whether the degree and pattern of variation in a fossil sample is consistent with the degrees and/or patterns of intraspecific variation seen in extant taxa that serve as analogs (i.e., reference samples or models) of intraspecific variation for fossil species. This is a null hypothesis testing approach.

The "null" hypothesis, by standard etymological and statistical definition, is a hypothesis of "no difference." Thus, it postulates that the nature of variation (both in degree and/or pattern) within a fossil sample is "no different" from that seen within a single species. Rejection of the null hypothesis leads to an alternate hypothesis that the fossil sample is more likely composed of multiple species.

The studies reviewed below have rejected the null hypothesis of intraspecific variation for the *H. habilis* sample. The main question reconsidered here is whether rejection of the null hypothesis is warranted in each case given a more rigorous and independent examination of the relevant data and/or methods of analyses.

The use of analog (reference) species

The criterion for rejecting the null hypothesis is whether the variation in the fossil sample is too great in degree of variation or too different in pattern of variation when compared to the most variable of the modern human and great ape analogs. This is a conservative approach, but is justified since there is no basis to assume a priori that the degree or pattern of variation in the *H. habilis* sample must be chimpanzee-like

rather than gorilla-like, or modern human-like. Indeed, *A. boisei,* which is contemporaneous with *H. habilis* and thus of equal proximity to the apparently unresolved human–chimpgorilla split (see Marks, 1993; 1994; Rogers, 1993; Deinard et al., 1998; Deinard and Kidd, 1999), is recognized as having a gorilla-like degree of sexual dimorphism (Leakey and Walker, 1988; Walker and Leakey, 1988). Also, *A. afarensis,* which is commonly considered to be ancestral to *H. habilis,* exhibits a level of dimorphism either intermediate between modern humans and gorillas (e.g., McHenry, 1991) or approximating gorillas and orangutans (e.g., Lockwood et al., 1996).

The inclusion of *Pongo* as a relevant analog is appropriate given the possibility that patterns of variation may be unique for each hominoid species (Oxnard, 1983a). Although some data suggests that gorillas and chimpanzees share a similar pattern of variation (Wood et al., 1991), more recent work supports the possibility of unique patterns among hominoid species (Uchida, 1992; Yaroch and Thompson, 1996). Early hominids, which are neither modern humans nor great apes, might still show other unique patterns of variation (e.g., see Oxnard, 1973, 1975, 1983b, 1987). Including *Pongo* thus broadens the analog base of large-bodied hominoids and allows the examination of another hominoid pattern of variation while, at the same time, not introducing a greater degree of variation than is seen among the African ape analogs.

Another consideration in analog choice is that intraspecific variation is a multifactorial phenomenon that is expressed differently across taxa. In the expression of sexual dimorphism, for example, other factors may have a greater impact on morphology than phylogenetic relationships (Oxnard, 1983a). These may include intrasexual competition, social organization, troop defense, food availability, or locomotor pattern among other factors. For a fossil species, all the various socioecological factors affecting the expression of morphological variation cannot be precisely known. Therefore, a broader comparative approach than simple phylogenetic similarity would suggest is justified. Moreover, to provide additional comparative perspective, an examination of a broader range of primates can be instructive.

The alternative is to recognize that the socioecological factors affecting variation such as sexual dimorphism are unique for each hominoid species and as such, there are no truly relevant analogs for early hominids. This, however, would prevent all scientific testing and reduce fossil analyses to mere speculation. So, one must proceed with the operational assumptions and extant analogs that are deemed reasonable and necessary. This is consistent with common practice (e.g., Stringer, 1986; Chamberlain, 1987; Lieberman et al., 1988; Wood, 1991, 1993; Kramer et al., 1995; Richmond and Jungers, 1995; Grine et al., 1996; Lockwood et al., 1996; among many others) and should need no further justification.

Although other hominid fossil taxa might be desirable as analogs due to their phylogenetic proximity to *H. habilis,* they are not used here because the homogeneity of no fossil hominid taxon is beyond dispute. The same controversy about species number plagues most fossil hominid taxa.

REVIEW AND REANALYSIS

Relative endocranial volume variation [the coefficient of variation (CV) of ECV]

The evidence from ECV variation for multiple species in *H. habilis* has centered on the use of the CV. Two arguments were advanced, each relying on a CV of 10 as an upper limit of variation for single species: (1) Stringer (1986) argued that a CV of 12.7 for *H. habilis* indicates a variability greater than that for any single species of fossil or extant hominoid taxa, which were represented as having CVs no greater than about 10; and (2) Wood (1985) used a CV of 10 to calculate a low probability that KNM-ER 1470 and KNM-ER 1813 are members of the same population. These two arguments were addressed by Miller (1991) who demonstrated that the CV of 12.7 for *H. habilis* actually falls within the range of published CV values for extant hominoids: *Pan troglodytes,* 9.7; *Gorilla gorilla,* 13.1; *Pongo pygmaeus,* 11.0 (great ape data from Tobias, 1971); modern humans, 6.1–13.5 with a

mean of 9.8 (data from Miller, 1991). Furthermore, Miller (1991) demonstrated that the 95% confidence limits for the *H. habilis* CV of 12.7 are so broad (5.1–20.3) as to make the CV unreliable for biologically meaningful interpretation. Contrary to previous claims (e.g., Wood, 1985; Stringer, 1986), the CV data provide no evidence of multiple species in the *H. habilis* sample.

Sexual dimorphism of endocranial volume

The question of "excessive" ECV variation can be explored in other ways, for example, by examining the level of sexual dimorphism. If the suspected level of dimorphism in *H. habilis* could be shown to be much greater than that for extant hominoids, then it might indicate a heterogeneous fossil sample. This possibility can be explored by examining the mean index of sexual dimorphism (mean Id) for ECV.

Gender of fossil specimens. The index can be determined in a variety of ways (e.g., see Tobias, 1975), but here it is calculated as the mean male ECV divided by the mean female ECV. Tobias (1987, 1991) suggested, on grounds other than ECV, that OH 7 (687 ml; ECV estimate from Holloway, 1978), OH 16 (667 ml; Stringer, 1986) and ER 1470 (752 ml; Holloway, 1978) are males and OH 13 (650 ml; Holloway, 1973), OH 24 (590 ml; Holloway, 1973), and ER 1813 (510 ml; Holloway, 1983) are females. If Tobias's determinations of gender for these specimens are accepted, and KNM-ER 1590 (750 ml; Stringer, 1986) and KNM-ER 1805 (582 ml; Holloway, 1978) are included as probable males, and KNM-ER 3732 (700 ml; Stringer, 1986) is included as a probable female (cf. Rightmire, 1993), then the mean Id is 1.12. Caution is needed since gender assignments may not be correct and the ECVs of fragmentary specimens (e.g., OH 7, OH 13, OH 16, KNM-ER 1590, KNM-ER 3732) are only approximations.

Comparison with all primates. With the above caveats in mind, it is useful to begin from a broad comparative perspective. Figure 1A shows the mean Id value of 1.12 for *H. habilis* compared to the mean index of sexual dimorphism for small samples

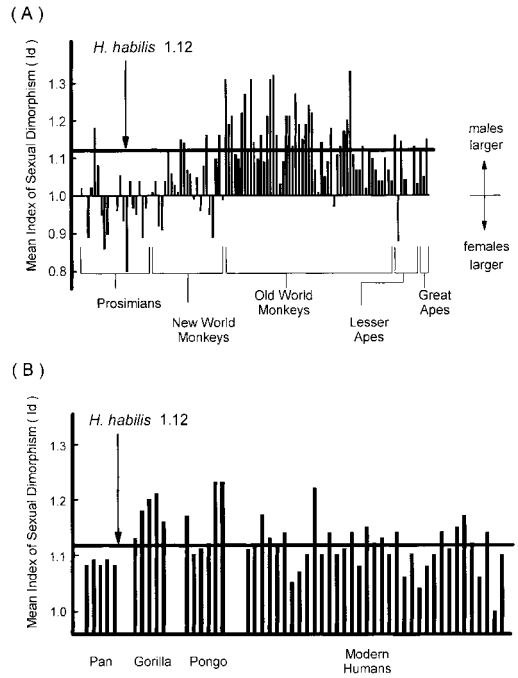

Fig. 1. (**A**) Mean index of sexual dimorphism (Id) of ECV for 108 primate species compared to *H. habilis.* (**B**) Mean index of sexual dimorphism (Id) for different samples of extant hominoids compared to *H. habilis* (see Appendix A for data).

($N \leq 6$) of 108 primate species calculated from unpublished data from Harvey and Clutton-Brock (personal communication; see Harvey and Clutton-Brock, 1985; Harvey et al., 1987). These data are particularly useful because they broadly sample the primates, but may not provide the best estimates of dimorphism possible for each species because of the small sample sizes involved.

The mean index of 1.12 for *H. habilis* is well within the range of mean indices for other primate species (0.81–1.33). Among the 108 primate species, 41 species (38%) have mean Ids of 1.12 or greater (mostly cercopithecoids). The following hominoid samples have mean Id's comparable to those for *H. habilis: Hylobates agilis* (1.16), *H. hoolock* (1.14), *H. pileatus* (1.13), *P. pygmaeus* (1.12), and *G. gorilla* (1.15). It is unlikely that gibbons actually have an ECV dimorphism greater than orangutans and gorillas; more likely, the Id values simply reflect poor estimates due to small samples.

Likewise, similar sampling error cannot be ruled out for the small *H. habilis* sample and its present calculated mean Id may itself be a poor estimate of the true population value for the species. Nevertheless, the present Id data, taken as a whole, indicate that the *H. habilis* sample, when compared to a broad range of primates, is not unusual in its level of ECV dimorphism.

Comparison with great apes. It is perhaps more relevant to compare the *H. habilis* Id with data for larger samples of modern large-bodied hominoids. Tobias (1971) compiled and published extensive ECV data for the great apes. From these data (see Appendix A), I calculated the mean Ids for each hominoid sample (Appendix A and Fig. 1B). The grand weighted mean Id's for each genus are: *Pan* (1.09), *Gorilla* (1.18), and *Pongo* (1.17; excluding the Oppenheim sample).

Although the *H. habilis* mean Id of 1.12 falls beyond the range of mean Id values for five samples of *Pan* (1.08–1.09), it is exceeded by the mean Id values for the five samples of *Gorilla* (1.13–1.21) and falls within the lower range of mean Id values for the six samples of *Pongo* (1.10–1.23), but is exceeded by the grand weighted mean Id for both *Gorilla* (1.18) and *Pongo* (1.17). Thus, when compared to large-bodied hominoids, the *H. habilis* sample is not unusual in its level of ECV dimorphism.

Comparison with modern humans. An examination of modern human Id data is equally relevant and instructive. The mean Id values for published male and female ECV data for 35 modern human samples range from 1.00 to 1.22 (Fig. 1B, see Appendix A). These samples are each geographically restricted and taken together represent a broad sampling of modern humans from all regions of the world. Seventeen samples (49%) have mean Id values greater than or equal to the *H. habilis* value of 1.12. The grand weighted mean Id for the 35 modern human samples (1.11) is nearly identical to the *H. habilis* mean Id. Thus, when compared to modern humans, the *H. habilis* sample is not unusual in its level of ECV dimorphism.

Probability estimates of craniofacial sexual dimorphism

Lieberman et al., (1988) used a sample of gorillas to construct a simple probability model for testing whether sexual dimorphism could explain the craniofacial differences between KNM-ER 1470 (presumed male) and KNM-ER 1813 (presumed female). For each of 27 craniofacial measurements, the index of sexual dimorphism (Id = male value/female value) for these two specimens was compared to a distribution of Id's derived from 20 male-female pairs randomly selected from an overall sample of 40 gorillas (20 males, 20 females). A probability of 5% ($p = 0.05$) was assigned to the 1470/1813 value if it fell at the limits of the gorilla distribution. A probability of less than 5% ($p < 0.05$) was assigned if the Id for 1470/1813 fell outside of the gorilla distribution. These probabilities were interpreted as representing the likelihood of sampling a gorilla pair that was as morphologically different as KNM-ER 1470 and KNM-ER 1813. Presumably, if the 1470/1813 pair were no more sexually dimorphic than gorillas, then by chance alone only 1 or 2 of the 27 Id's ($\sim$5%) for the hominids should fall outside the gorilla distributions.

Lieberman et al. (1988), however, found that 11 out of 27 (41%) of the Ids for 1470/1813 fell at or beyond the limits of the gorilla distributions. Under the a priori assumption that *H. habilis,* as represented by KNM-ER 1470 and KNM-ER 1813, was unlikely to be more dimorphic than gorillas, these results were interpreted as indicating that the two fossils belong in different taxa. However, these results and interpretation are problematic on a number of grounds. The first is that their method is untested on samples of known taxonomic composition to verify its effectiveness as a discriminator between intraspecific and interspecific variation. The second is that the method runs into statistical problems which are treated below.

A counter-example. A simple test of the method of Lieberman et al. (1988) is to treat known hominoid specimens of the same species as "fossils" and subject them to the

same gorilla standard and interpretive analysis as was done for the two *H. habilis* fossils. Thus, I measured two male and two female *P. pygmaeus* crania at the Museum of Comparative Zoology (Harvard) for 24 of the 27 measurements used in the Lieberman et al. (1988) study (three nonstandard measurements could not be duplicated). The four specimens measured are wild-caught animals from a single locality in northern Sumatra belonging to a single subspecies, *P. p. abelii.* These specimens were not preselected for extreme dimorphism, nor is it likely that they represent the limits of dimorphism of the subspecies or species as a whole. They provide a good test case since they represent one subspecies of a single hominoid species that is about as dimorphic as *G. gorilla.*

Rather than presenting the results for a single pair, the Ids for the four possible male-female pairs (A–D) were calculated and compared to Lieberman et al.'s (1988) gorilla distributions. For pair A, 12 out of 24 or 50% of the Ids are at or beyond the 95% population limits of the gorilla distributions. The number of significant comparisons for the other *P. p. abelii* pairs are as follows: pair B (67%), pair C (46%), and pair D (38%). These results encompass and exceed those for the KNM-ER 1470/KNM-ER 1813 comparison (41%). These results show clearly that the probabilistic method of Lieberman et al. (1988) does not discriminate between intraspecific variation and interspecific variation.

Statistical problems. The other problem with the method of Lieberman et al. (1988) is that it is equivalent to conducting 27 separate *t* tests, each at a significance level (α) of 5%. This is acceptable as long as the tests are independent of one another. However, when the tests are not independent, the significance levels are affected such that the probability of committing a type I error (rejecting a null hypothesis that is, in fact, true) increases with the number of nonindependent tests or comparisons. By the 20th nonindependent comparison, the probability is 92% that a type I error will be made if each of the tests is conducted at a 5% significance level (Zar, 1984). Nonindependent *t* tests require a correction so that the overall experiment-wise significance level remains at 5%. An approximate solution to obtain the proper significance level for each test is to divide the overall significance level by the number of nonindependent tests.

The intercorrelations of craniofacial traits pose just this problem of nonindependent comparisons for the Lieberman et al. (1988) analysis. Applying the rough correction suggested above reduces the significance level from 5% to 0.2% for each of the 27 tests by Lieberman et al. (1988). Thus, none of the *H. habilis* Ids are statistically significant unless they could be shown to have a probability of less than 0.2% for each test. However, the manner in which probabilities were assigned in the analysis by Lieberman et al. (1988) does not allow for an independent assessment of this.

Craniofacial characters are generally intercorrelated. For example, Wood (1975) found that 14–34% of skull characters exhibited statistically significant intercorrelations in same sex samples of living hominoids. These percentages would probably be higher if size dimorphic males and females were combined in the same analysis (as might be found in a mixed-sex fossil sample).

Correlation analysis. To further investigate the question of intercorrelation of craniofacial characters raised above, I conducted a correlation analysis for 25 male and 35 female wild-caught specimens of *P. p. pygmaeus* (measured at the National Museum of Natural History, Smithsonian). Correlations between the following craniofacial characters were determined: glabella-prosthion, outer M^2 width, nasion-rhinion, orbital height, orbital breadth, post-orbital constriction, and maximum biparietal width. These are seven of the eleven characters found to be significant ($P \leq 0.05$) by Lieberman et al. (1988). Male and females were first analyzed separately and then together.

The increased correlation for pooled-sex samples can be seen in the following examples. For males, the correlations between glabella-prosthion and outer M^2 width, outer M^2 width and nasion-rhinion, nasion-

rhinion and orbital breath, and orbital height and maximum biparietal width are 0.31, 0.31, 0.21, and 0.04, respectively. For females, the correlations are 0.60, 0.33, 0.18, and 0.27, respectively. When both sexes are analyzed together, the correlations become 0.80, 0.64, 0.54, 0.39, respectively.

Intercorrelations among craniofacial characters are greater in mixed-sex samples that encompass a greater size range than is found in either sex alone. So for mixed-sex fossil samples, the problem of intercorrelation of craniofacial characters is even greater than Wood's (1975) data suggest. Such potentially high intercorrelations among characters include those variables that were singled out as being significant in the Lieberman et al. (1988) analysis. Because the simple *t* testing approach of Lieberman et al. (1988) failed to adjust for correlated characters, it provides no statistical or probabilistic basis for excluding KNM-ER 1470 and KNM-ER 1813 from being in the same taxon.

The question of size

Perhaps the "significant" differences, found by Lieberman et al. (1988), were simply a reflection of the obvious difference in size between KNM-ER 1470 and KNM-ER 1813. This raises the question of whether the size difference between the two specimens is too great to be accommodated within a single species. In answering this question, the problem of correlated characters can be avoided by using principal components analysis (PCA) to summarize multivariate data in a way that allows the within-group structure to be examined while preserving the relationships among the specimens. This analytical technique allows variation among fossil specimens to be compared visually to the within-group variation of modern analog species.

To investigate the size difference between KNM-ER 1470 and KNM-ER 1813 in a preliminary way, I measured 25 male and 27 female, wild-caught specimens of *P. p. pygmaeus* from the National Museum of Natural History (Smithsonian) for eight craniofacial traits that broadly reflect the overall size of the cranium. These measurements

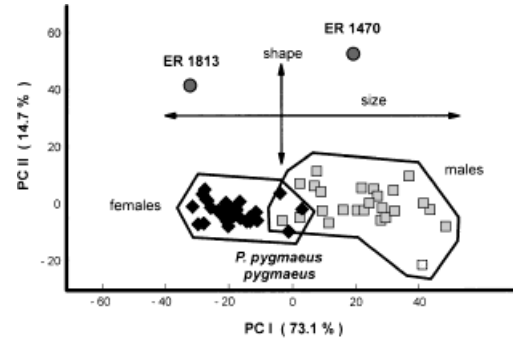

Fig. 2. Principal component analysis of KNM-ER 1470 and KNM-ER 1813 compared to the within-group variation of *P. p. pygmaeus* (25 males, 27 females). PC I (reflecting size and size-related shape) shows that the distance between KNM-ER 1470 and KNM-ER 1813 is similar to that separating average male and female Pongo. PC II (reflecting size-independent shape) shows that the distance between the fossils is easily matched by within-sex or between-sex distances in the *Pongo* sample.

are: (1) upper facial height, (2) biorbital breadth, (3) bimaxillary breadth, (4) maximum cranial length, (5) biporionic breadth, (6) post-orbital constriction, (7) outer M^2 breadth, and (8) malar height. Measurements of the fossil hominids were taken from the literature (Wood, 1991).

A PCA of the raw data, using a variance-covariance matrix, was then applied to all the data. When the individual scores for the first two PC axes are plotted (87.8% of the variation), it is clear that the differences between KNM-ER 1470 and KNM-ER 1813 are consistent with variation in the moderately-sized sample of *P. p. pygmaeus* (Fig. 2).

The first PC axis reflects the size and size-related shape differences between sexes and between specimens of the same sex. The smallest female specimens are positioned at the extreme left and the largest male specimens are located at the extreme right. The distance between KNM-ER 1470 and KNM-ER 1813 is a little greater than that between average male and female *P. p. pygmaeus* specimens. However, it is less than the distance between many male and female pairs of the *P. p. pygmaeus* sample. Interestingly, the interfossil distance is no greater than that between the largest and smallest male specimens of *P. p. pygmaeus.*

If size and size-related shape is largely reflected in the first PC axis, then the sec-

ond axis represents size-independent shape information uncorrelated with the first axis. Here again, the distance between the fossils on the second axis is consistent with distances within or between the sexes of the *P. p. pygmaeus* sample.

These results indicate that the overall size (including size-related shape) difference between KNM-ER 1470 and KNM-ER 1813 does not constitute evidence that they must belong in different taxa. These results also suggest that the magnitude of size-independent shape difference between the fossils is no different than that among members of the same sex (male or female) within a single large-bodied hominoid subspecies exhibiting *Pongo*-like variation. Using an increased number of craniofacial variables, these results have been substantiated for *G. gorilla* (Miller, 1995; Miller et al., 1997), *P. pygmaeus* (Miller, 1995), and modern *H. sapiens* (Miller et al., 1998).

Multivariate analysis of the facial region

It has been argued that distinctive facial differences between KNM-ER 1470 and KNM-ER 1813 rule out conspecificity. Wood (1985), citing data from a multivariate analysis of the facial region of hominids later published in full by Bilsborough and Wood (1988), observed that the generalized distance (D) of 3.4 between KNM-ER 1470 and KNM-ER 1813 was greater than the distances of 2.8 between *A. africanus* and KNM-ER 1813, 3.3 between *A. boisei* and KNM-ER 1470, and 3.1 between Neandertals and modern humans. This was interpreted as evidence that KNM-ER 1470 and KNM-ER 1813 may not represent the same taxon.

However, when the full generalized distance matrix of Bilsborough and Wood (1988) is examined, the above interpretation becomes less tenable. The full matrix shows interspecific distances such as 2.7 between *A. africanus* and *A. robustus* and 4.3 between *H. erectus* and *A. boisei*. However, it also shows an overlapping range of intraspecific distances such as 2.5 between OH 24 and KNM-ER 1813 (whom few would doubt are conspecifics) and 3.8 between KNM-ER 732 and KNM-ER 406 (presumptive female and male *A. boisei* specimens). In this context, the value of 3.4 for *H. habilis* cannot be interpreted as excluding the null hypothesis of intraspecific variation. Indeed, it is remarkable that some investigators are apparently not willing to accept a degree of variation in the *H. habilis* sample that they accept in other hominid taxa such as *A. boisei* (e.g., Wood, 1991, 1993; Wood et al., 1994; among others).

Cranial angle variation

Stringer (1986) argued that the overall degree and pattern of cranial angle variation indicate that *H. habilis* specimens KNM-ER 1470, KNM-ER 1813, and OH 24 represented multiple species. He noted that KNM-ER 1470 and OH 24, or KNM-ER 1470 and KNM-ER 1813 might be accommodated in a single species, but not all three. Additionally, he claimed that the particular pattern of differences between KNM-ER 1470 (presumptive male) and KNM-ER 1813 (presumptive female) for some cranial angles (the SSA in particular) is inconsistent with sexual dimorphism. Thus, in reanalyzing the cranial angle data, there are two questions to address: (1) is the degree of cranial angle variation "excessive" in the *H. habilis* sample? and (2) is the pattern of variation inconsistent with sexual dimorphism or other forms of intraspecific variation?

The cranial angles fall into three groups. The first group comprises the angles of the facial triangle including the prosthion angle (PRA), nasion angle (NAA), and basion angle (BAA) (Fig. 3A). The second group comprises the angles of the anterior cranial triangle including the bregma angle (BRA), nasion-bregma angle (NBA), and basion-bregma angle (BBA) (Fig. 3A). The third group comprises the angles of two transverse facial angles including the subspinale angle (SSA) and nasiofrontal angle (NFA) (Fig. 3B). These angles were determined trigonometrically by Stringer (1986) from distances measured between the appropriate osteometric landmarks on casts of the fossils.

Degree of variation. An appropriate measure is required to assess whether cranial angle variability is "excessive." The

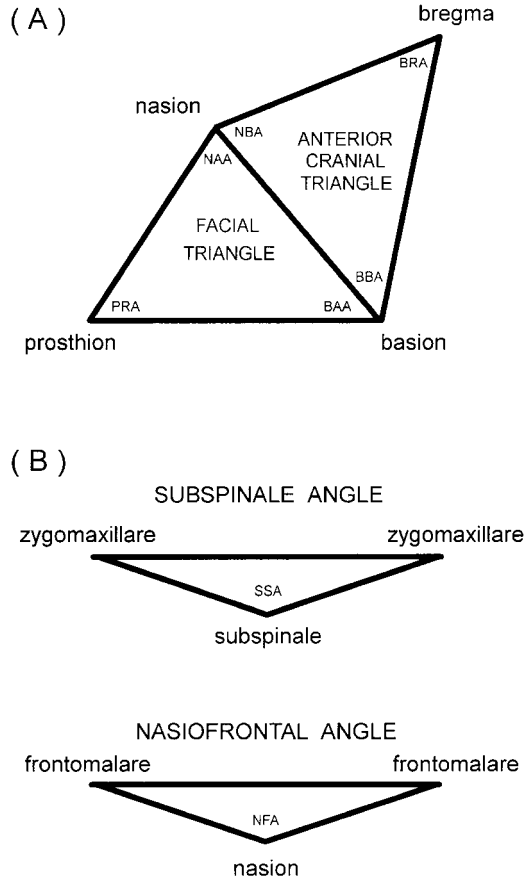

Fig. 3. Cranial angles between the osteometric landmarks used in Stringer's (1986) study. (**A**) Angles of the facial triangle (NAA, PRA, BAA) and anterior cranial triangle (NBA, BRA, BBA). (**B**) Transverse facial angles (SSA, NFA).

small size of the *H. habilis* sample (KNM-ER 1470, KNM-ER 1813, and OH 24) precludes the traditional use of CV's (for a discussion of the use of the CV in fossil species recognition studies see Cope, 1993; and Kelley and Plavcan, 1998). An alternative is to determine if the observed range for the *H. habilis* sample exceeds the 95% population limits (calculated as the mean $\pm$ 2 SD) expected for single hominoid species.

This comparison of ranges is essentially what Stringer does in his bivariate plots (of PRA versus NAA, NBA versus NAA, and NFA versus SSA), when he compares the values for the three *H. habilis* specimens to the 95% population limits and mean values for *P. troglodytes* and modern humans (see Stringer, 1986, Figs. 5–7). This kind of comparison, though imperfect and not a strict statistical test, is nevertheless a reasonable way to judge excessive variation in this case. Certainly, if the observed range for the *H. habilis* sample greatly exceeded the 95% population limits for modern hominoid species, then an interpretation of excessive variation might be warranted.

Stringer (1986) presented cranial angle data for 1,652 modern humans, 16 *P. troglodytes,* 29 *G. gorilla,* 28 *P. pygmaeus,* and 3 *H. habilis* specimens. The observed range for *H. habilis* and the 95% population limits (a.k.a. normal range) for Stringer's (1986) comparative hominoid samples appear in Appendix B. In seven out of eight cases, Stringer's (1986) values for the observed range for the *H. habilis* sample are not excessive compared to the 95% population limits of one or more hominoid analogs. In particular, for all the angles of the facial and anterior cranial triangles, the *H. habilis* range does not exceed the 95% population limits for *G. gorilla.* In four out of six angles, the *H. habilis* range is less than the limits for *P. pygmaeus.* For the two transverse facial angles, only the observed range for the *H. habilis* NFA is excessive compared to the hominoid analogs. However, a note of caution is required in interpreting this result.

New measurements. Using the same trigonometric protocol as Stringer (1986), except where noted below, I used my own linear measurements on casts of the fossil specimens to conduct an independent assessment of cranial angle determination. A different method was used to measure the NFA since: (1) this was the most variable of the angles measured by Stringer (1986), and (2) there is greater measurement error for broad angles that are trigonometrically determined (see Trigonometric Problems below). Consequently, a direct protractor measurement was used to estimate the NFA. My values for all angles appear alongside those of Stringer's in Appendix B.

My measurements show that the observed range for five out of the eight cranial angles (NAA, BAA, PRA, BBA, and SSA) for the *H. habilis* sample are less than or equal to the 95% population limits for all modern

hominoids. For all angles, the variation in the *H. habilis* sample is less than the 95% population limits for *G. gorilla* and *P. pygmaeus.* With respect to the NFA, the observed range of 17° for the *H. habilis* sample is greater than the 95% population limits for modern humans (10°), but less than the 95% population limits for *P. troglodytes* (18°), *G. gorilla* (22°), and *P. pygmaeus* (20°).

Inter-observer error. The differences between Stringer's (1986) measurements and mine are sometimes small and sometimes large (see Appendix B). There is no particular reason to believe that my measurements are any better than his, or vice versa. However, my measurements for the NFA agree more with Rightmire's (1993) angle calculations that were based on measurements taken on the original specimens. For Rightmire's (1993) data, the observed range for the NFA is 12°, which is slightly greater than the 95% population limits for modern humans, but less than that for all the great apes.

The inter-observer differences in angular measurements require comment. Such differences are likely due to a number of factors: (1) cast variability, (2) inter-observer error in locating osteometric landmarks, and (3) problems inherent in the trigonometric calculation of angles.

The first factor deserves more attention since the practice of measuring casts in paleoanthropological studies is not uncommon. Stringer (1986) noticed that his measurements on the casts differed from published measurements on the original fossils by as much as 5%. It is uncertain whether such differences reflect physical inaccuracies in casting (e.g., shrinkage) or differences in measurement technique.

With regard to the second factor, workers may vary in locating the position of osteometric landmarks that are present and observable in the original fossil specimen. Greater differences may be expected when landmarks are present but not easily identifiable on a cast (e.g., due to poor cast quality), displaced (e.g., due to distorted morphology), or estimated because of missing bone or poor preservation of anatomical detail. These factors surely pose problems in the measurement of casts representing fragmentary, plastically distorted, reconstructed fossil specimens.

The problems just mentioned may occur, for example, when measuring the NFA for OH 24. To calculate this angle trigonometrically, it is necessary to measure between three points: nasion and right and left frontomalare. On OH 24, the left frontomalare point is completely missing, which affects two of three dimensions used to calculate the NFA. Additionally, given the extensive reconstruction of this specimen (see Reader, 1981 for "before and after" photographs; Tobias, 1991, plates 1–3), it is likely that the right frontomalare and nasion points are displaced relative to one another. A similar situation exists for the SSA where one of three landmarks are missing in OH 24 and KNM-ER 1813, and all three are missing in KNM-ER 1470. Given this problem of missing and displaced landmarks, the taxonomic value of estimating these cranial angles is highly questionable.

There is the additional problem for KNM-ER 1470 of whether the face is set on the cranium at the biologically correct angle (Bromage, 1992, 1993; Grine et al., 1996). There is only one meager contact in the region of nasion, and even this "articulation" is questionable. In view of this uncertainty, which affects all three angles of the facial triangle for KNM-ER 1470, calculated angles for this triangle must be regarded as highly suspect.

Trigonometric problems. The problems mentioned above pose a particular difficulty for trigonometric determinations of cranial angles because a difference of 1 or 2 mm in a measurement can result in a substantial difference in the calculated angle. For example, if the frontomalare-frontomalare measurement is 83 mm and the frontomalare-nasion (fm-n) measurement on both sides is 42 mm, then the NFA is 162.3°. However, if the fm-n measurements are slightly reduced to 41.5 mm, then the NFA changes by 12.1° to 174.4°.

The confidence with which angles can be interpreted is seriously undermined when small changes in linear measurements (comparable to measurement error) result

in substantial changes in derived angles. This is a critical problem for fossils when landmarks must be estimated or are distorted in position due to fossilization and/or reconstructive processes. When all the possibilities for nonbiological variation are taken into account, the degree of cranial angle variation among KNM-ER 1470, KNM-ER 1813, and OH 24 does not provide a compelling basis for rejecting the null hypothesis of intraspecific variation.

Pattern of variation. Stringer (1986, p. 287) stated that sexual dimorphism could explain differences in SSA between KNM-ER 1470 and KNM-ER 1813 "only if it were argued that Plio-Pleistocene hominids had an atypical pattern of dimorphism in lower facial projection." He observed that SSA values for the great apes are generally smaller in males than in females, whereas the SSA for KNM-ER 1470 (presumed male) is larger than for KNM-ER 1813 (presumed female). Rightmire (1993) also makes this claim for the SSA as well as the NFA based, presumably, on Stringer's comparative data (1986, Table 6).

However, *t* tests that I conducted on Stinger's (1986, Table 6) data indicate that only *P. pygmaeus* among hominoids shows a significant sexual difference ($P < 0.001$) for the SSA, with males having smaller SSA values (mean = 108.4) than females (mean = 117.2) on average. Furthermore, the 95% population limits for males and females overlap completely in modern *H. sapiens, P. troglodytes,* and *G. gorilla,* and overlap substantially in *P. pygmaeus.* Similar results are obtained for the NFA for which only modern humans show a statistically significant difference between males and females ($P < 0.001$). But this difference is extremely small (0.7°) and the 95% population limits for the two sexes overlap completely. Thus, for both the NFA and SSA, there is substantial overlap between the sexes. Consequently, no pattern of dimorphism can be claimed as "typical" for great apes or modern humans. The cranial angle data, then, provide no basis, either from degree or pattern of variation, to reject the null hypothesis of intraspecific variation for the *H. habilis* sample.

CV analyses of craniofacial variation

Wood (1991, 1993). Another attempt at assessing the degree of craniofacial variation in *H. habilis* was made by Wood (1991, 1993) who used the CV to assess variation in 19 craniofacial measurements. He compared the CVs for the *H. habilis* sample to those for reference samples of *G. g. gorilla, A. boisei,* and *H. erectus.* In every case, the 95% confidence limits for the *H. habilis* CV encompasses the CVs of the reference samples. Wood (1991, p. 240, 1993, p. 494) acknowledged that "the observed degree of variation within the overall early *Homo* hypodigm suggests that few of the measurements considered are so variable that a single taxon solution is out of the question." In other words, the null hypothesis cannot be rejected on the basis of these data.

If the above CV comparison were treated as a formal statistical test, it would suffer from the same statistical problems of intercorrelated cranial traits as reviewed above for Lieberman et al. (1988). However, even informal inspection of the extremely broad confidence limits of the *H. habilis* CV's (Wood, 1993, Table 8) reveals they are not biologically meaningful. The problem is extremely small sample sizes which, except for cranial capacity (N = 9), range from 3 to 5 specimens.

Kramer et al. (1995). Kramer et al. (1995) did a similar CV analysis of 14 craniofacial variables. However, due to the varying incompleteness of fossil specimens, their sample size for each variable ranged from 2 to 7, with an average sample size of 3.6. For each variable, they compared the CVs of the *H. habilis* sample to a randomized distribution derived by repetitively drawing like-sized numbers of specimens from moderate samples of modern humans, *P. troglodytes, G. gorilla,* and *P. pygmaeus* (999 trials for each variable for each analog species).

They found for 9 of 14 variables (64%) that the CVs for the *H. habilis* sample were greater than the 95% limits for modern humans and *P. troglodytes,* but only three CVs were "significantly" different relative to *P. pygmaeus* and only two relative to *G. gorilla* (Kramer et al., 1995). However, realizing

the problem of conducting nonindependent multiple *t* tests, they did not use this data as the basis for testing the single species hypothesis but rather relied on their variability (CV) profile data to reject that hypothesis (see Variability Profiles below). So, the CV data presented in these studies (Wood, 1991, 1993; Kramer, 1995), as noted by the authors themselves, do not provide a basis to reject the null hypothesis of intraspecific variation for the *H. habilis* sample.

Multivariate pattern of sexual dimorphism

Wood (1991, 1993) suggested that the clearest evidence for multiple species in *H. habilis* comes from the multivariate pattern of sexual dimorphism. In those studies, the results of two canonical variate analyses (CVA) were presented. One CVA, based on 16 cranial variables, compared KNM-ER 1470 and KNM-ER 1813 to the canonical variate means of male and female samples of *P. troglodytes, P. pygmaeus, G. gorilla,* and modern humans. The other CVA, based on 14 cranial variables, compared KNM-ER 1470 and OH 24 to the canonical variate means of the same hominoid samples. The results of these analyses were similar in that the lines connecting ER 1470 (presumptive male) to either KNM-ER 1813 or OH 24 (presumptive females) are oriented in different directions when compared to the between-sex axes for *P. troglodytes, G. gorilla,* and modern humans. This "difference" was interpreted as strong evidence that the cranial differences between the three fossil specimens could not be explained by sexual dimorphism within a single species.

Wood's (1991: Fig 6.9; 1993: Fig. 5) CVA results for KNM-ER 1470 and KNM-ER 1813 are shown in Figure 4A. One problem with Wood's analysis is that two individual specimens, KNM-ER 1470 and KNM-ER 1813, were compared to male and female group means (centroids) for the hominoid analogs. The meaningfulness of comparing the orientation of the axis connecting two individuals with the axis connecting two group means is unclear without considerable statistical validation. Such comparisons are valid only if KNM-ER 1470 and KNM-ER 1813 represent the exact male and female means, respectively, for *H. habilis.* This, however, is difficult to sustain since these two fossils represent the morphological extremes of the present *H. habilis* sample. Other, albeit, more fragmentary specimens such as OH 7 (presumptive male), OH 13 (presumptive female), and OH 16 (presumptive male) are intermediate in size. Consequently, the male and female means for the *H. habilis* sample are likely to be less extreme than KNM-ER 1470 and KNM-ER 1813 suggest.

Another problem relates to how within-group, individual variation surrounding the male and female means of the hominoid analogs affects interpretations of the meaning of the distance and orientation of the fossil specimens to one another. In order to interpret the meaningfulness of an axis joining two fossil specimens, it is first necessary to understand the variation in axis orientation among male-female pairs of individuals from the analog samples of recent hominoids. Although Wood (1991, 1993) plotted only group centroids in the original graphs, the possible range of variation in axis directions among pairs of individual specimens can be demonstrated by plotting 95% equiprobability ellipses of within-group variation (i.e., 95% population limits). Figure 4B shows the pooled 95% population limits of within-group variation for a single sex plotted and centered on the midpoint of the line connecting KNM-ER 1470 and KNM-ER 1813. This ellipsoid is plotted as 2.0 standard deviations in radius from the group centroid and is scaled using the canonical variate axes, which are in standard deviation units.

An examination of these 95% population limits reveals that KNM-ER 1470 and KNM-ER 1813 are well inside the within-group variation for a single sex of the hominoid analogs (Fig. 4B). Furthermore, since this ellipsoid represents a "cloud" of individuals, it is easy to visualize a myriad of axes connecting all possible pairs of individuals. The plethora of line orientations connecting all possible pairs of individuals would be even more dramatic if one cloud was superimposed over KNM-ER 1470 and another over KNM-ER 1813, as if these specimens represented the male and female means for the *H. habilis* sample. Clearly, it would not

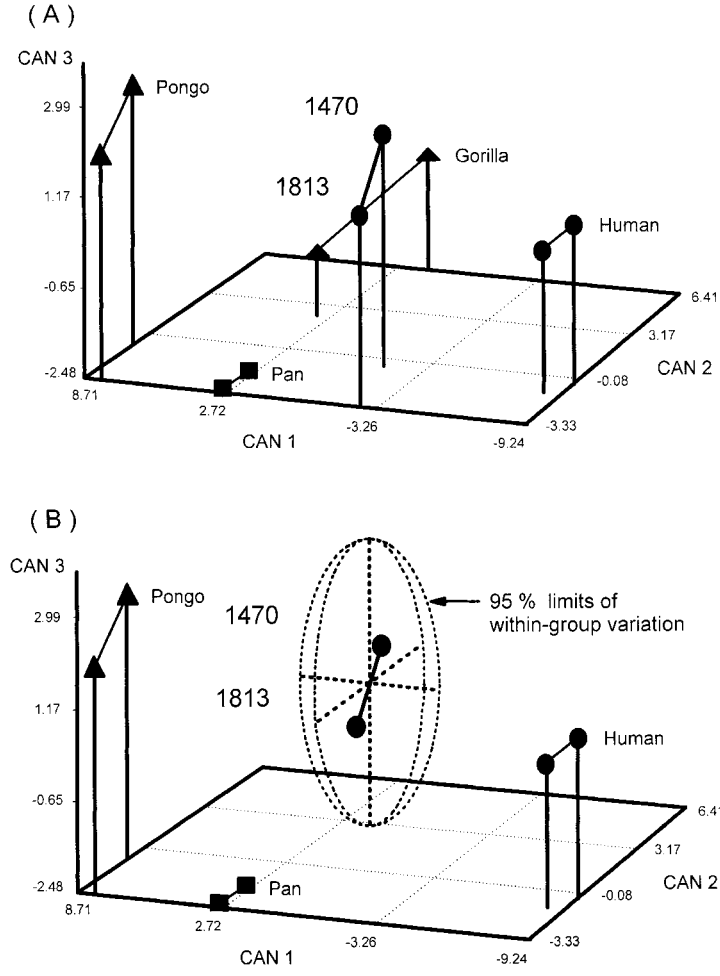

Fig. 4. (**A**) Wood's (1991, 1993) canonical variates analysis (CVA) plot for KNM-ER 1470 and KNM-ER 1813 compared to great apes and modern humans. (**B**) Wood's (1991, 1993) plot modified by including the within-group variation for a single sex. The 95% equiprobability ellipse (calculated as ±2 SD) is positioned at the midpoint of the line connecting KNM-ER 1470 and KNM-ER 1813. The dimensions of this ellipse are taken directly from the canonical variate axes, which are in standard deviation units.

be unusual to find two individuals whose axis differs as much as does the 1470/1813 axis from the axis connecting the male and female group centroids. This indicates that the orientation of the axis connecting any two particular individuals, such as KNM-ER 1470 and KNM-ER 1813, does not necessarily say anything about the pattern of sexual dimorphism in the species as a whole—especially since the difference between the two fossils is contained in an ellipsoid representing the within-group variation of a single sex. The CVA pattern of variation for KNM-ER 1470 and KNM-ER 1813 provides no basis to reject the null hypothesis of intraspecific variation in the *H. habilis* sample.

The results for Wood's (1991: Fig. 6.6; 1993: Fig. 4) CVA for KNM-ER 1470 and OH 24 are more enigmatic. The distance between KNM-ER 1470 and OH 24 is much greater than between KNM-ER 1470 and KNM-ER 1813. This is difficult to understand given the great similarity between OH 24 and KNM-ER 1813 that is evident when the raw measurements for these specimens are compared (e.g., see measurements in Chamberlain, 1987, or Wood, 1991). Given this similarity, the distances between each of the latter two and KNM-ER 1470 should be more similar than Wood's (1991, 1993) analysis indicate. Thus, the greater distance between OH 24 and KNM-ER 1470 is probably artifactual. One

possible explanation is that the results are affected by the distortion remaining in OH 24 after its substantial reconstruction. [See Reader (1991) for before and after photographs, Leakey et al. (1971) and Tobias (1991) for a description of the distortion remaining in OH 24.] Other explanations are also possible including, perhaps, the way that the fossils were methodologically interpolated into the multivariate analysis (see Albrecht, 1992). Nevertheless, given the great similarity of KNM-ER 1813 and OH 24 and the considerable distortion remaining in the latter, the disparity of Wood's (1991, 1993) CVA results when these specimens are compared separately to KNM-ER 1470 should be viewed with much caution until such a disparity can be independently verified.

Variability (CV) profiles

On the basis of variability profiles, Kramer et al. (1995) concluded that the *H. habilis* sample must represent a mixed-species sample because its CV profile was statistically different from those of all other extant large-bodied hominoids. A variability profile is simply a line graph of CV values for a number of variables for a given species or sample (Sokal and Braumann, 1980). Profiles for different taxa may be compared visually or statistically. Kramer et al. (1995) used a nonparametric test—Kendal's tau rank-order test—to statistically evaluate the CV profile for *H. habilis.* In this test, the rank order of the CVs (from smallest to largest) for 14 craniometric variables for the *H. habilis* CV profile was compared to the rank order of CVs in the profiles of hominoid referent samples.

The CVs comprising the *H. habilis* profile were determined by Kramer et al. (1995) on the basis of very small samples sizes (N = 2 to 7), which were different for each variable. Attempting to address the problem of small sample size, Kramer et al. (1995) used a randomization approach to construct the CV profiles for their referent samples. From each of their parent samples of *P. troglodytes, P. pygmaeus, G. gorilla,* and *H. sapiens,* they randomly drew 999 samples for each variable, of sample size equal to that for *H. habilis.* From these 999 samples, a mean CV was determined for each variable for each species, and these CV's were used to plot a variability profile for each analog species (Fig. 5A).

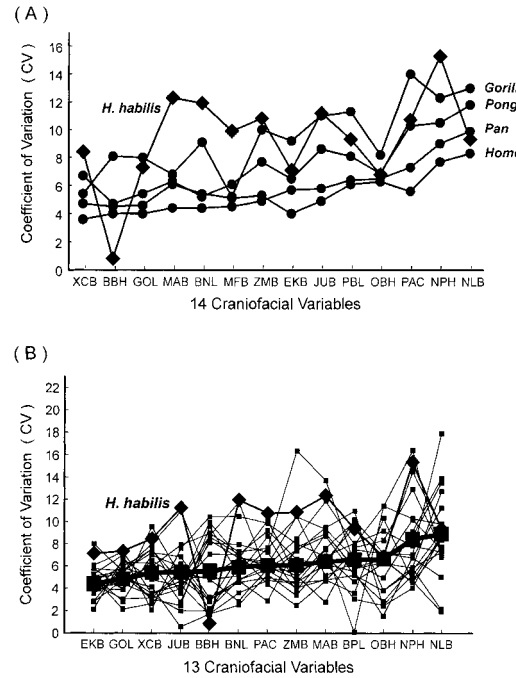

Fig. 5. Variability profiles. (**A**) Variability profiles for *H. habilis* (line with diamonds) compared to *G. gorilla, P. pygmaeus, P. troglodytes,* and modern humans (lines with black circles). (From Kramer et al., 1995.) (**B**) A counter-example comparing the variability profiles for 20 random small samples (N = 4, lines with small squares) of modern humans to the CV profile for the entire sample (heavier line with large squares).

The rank order of the CVs in the variability profile for the *H. habilis* sample was then tested against the rank order of CVs for the mean profiles of the hominoid referents using Kendal's tau rank-order test. Since the mean CV profiles for the ape analogs were found to be statistically similar to one another, any hominoid sample with a different CV profile was thought to represent multiple species. This assumption was made even though the profiles for modern *H. sapiens* and *G. gorilla* were not the same. Kramer et al. (1995) reported that *H. habilis* has a variability profile statistically different from all the referent samples. Consequently, the null hypothesis of intraspecific variation for the *H. habilis* sample was rejected.

The fallacy of this approach is in treating the CV profile for the very small *H. habilis* sample as equivalent to a mean "population" profile derived from a much larger sample as, indeed, was the case for the analog taxa. It is well known that statistics for small samples fluctuate widely and thus provide poor estimates of population values. Likewise, CVs of small samples have notoriously wide confidence limits and resultant CV profiles will be poor estimates of the mean population profile. This is shown in the following example adapted from Albrecht and Miller (1997 and unpublished data).

Howells' (personal communication, 1973, 1989, 1995) data set for modern humans (N = 2,524) was used to draw 20 random small samples of four specimens each. CV profiles for these small samples were constructed and treated as though they were the variability profiles of unknown "fossil hominid" samples. These profiles were then tested, with Kendal's tau rank-order test, against the mean profile for the total human sample. The results indicate that 14 of 20 (70%) randomly drawn small samples of humans had statistically different CV profiles than the mean profile for the entire human sample (Fig. 5B). Albrecht and Miller (1997) found similar results for *P. troglodytes* (55%), *G. gorilla* (45%), and *P. pygmaeus* (30%). We concluded that this method of using CV profiles cannot be used reliably on small fossil samples to make decisions about whether such samples contain more than one species. Consequently, CV profiles provide no basis to reject the null hypothesis of intraspecific variation in the *H. habilis* sample.

Distance analysis using exact randomization procedures

Several studies involving distance data and randomization methods presented similar results and conclusions regarding *H. habilis* (Kramer et al., 1995; Grine et al., 1996; and Donnelly, 1996). These studies all rejected the null hypothesis, yet their data suggest otherwise.

Principal component distances. Kramer et al. (1995) used a randomization approach to determine if KNM-ER 1470, KNM-ER 1813, and OH 24 represent the same species. Principal component (PC) distances, based on 14 craniofacial measurements, among pairs of specimens were compared to distributions of similar pair-wise distances for analog samples of 120 modern *H. sapiens,* 62 *P. troglodytes,* 29 *P. pygmaeus,* and 61 *G. gorilla.* For the apes, the distribution of PC distances resulted from exact randomization based on all possible pairs of each species (male-male, female-female, and male-female). For modern *H. sapiens,* the distribution was based on 2,499 distances between randomly selected pairs of individuals. The PC distance between each pair of specimens was calculated by squaring the difference in their PC scores for each PC axis, weighting the squared differences by the eigenvalue for that PC axis, and then adding these distances for all axes to arrive at the total "PC" distance. Presumably, the idea behind weighting the distances by the eigenvalues was to give greater weight to those PC axes that account for greater amounts of the variation.

The probability of drawing a pair from the analog sample as different as the fossil pair was determined by comparing the PC distance between a pair of fossils to the distribution of pair-wise distances for each analog sample. The results for KNM-ER 1470 and KNM-ER 1813 compared to each analog species were: modern *H. sapiens* ($P < 0.002$), *P. troglodytes* ($P < 0.035$), *P. pygmaeus* ($P < 0.103$), and *G. gorilla* ($P < 0.220$) (Fig. 6). In other words, only 0.2% of the modern *H. sapiens* pairs had PC distances greater than or equal that for the pair of fossils, whereas 22% of the *G. gorilla* pairs exceeded the fossil distance. The results for KNM-ER 1470 and OH 24 were: modern *H. sapiens* ($P < 0.046$), *P. troglodytes* ($P < 0.087$), *P. pygmaeus* ($P < 0.261$), and *G. gorilla* ($P < 0.405$) (Fig. 6). This means that from 4.6% of the modern *H. sapiens* pairs to 40.5% of the *G. gorilla* pairs exceeded the distance between the fossil specimens. Thus, at a 5% probability level, the KNM-ER 1470/1813 distance was not significant compared to *P. pygmaeus* and *G. gorilla,* and the KNM-ER 1470/OH 24 distance was not significant when compared to any ape species.

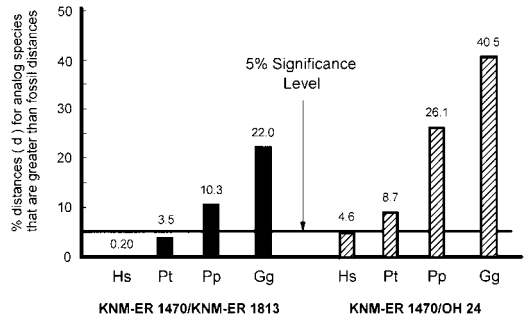

Fig. 6. Randomization data of Kramer et al. (1995) based on 14 variables. The vertical bars represent the percentages of weighted principal component distances among 2,499 randomized pairs of modern humans (Hs), and among all possible pairs of *P. troglodytes* (Pt), *P. pygmaeus* (Pp), and *G. gorilla* (Gg) that are greater than distances between KNM-ER 1470 and KNM-ER 1813 (**left**) or between KNM-ER 1470 and OH 24 (**right**).

Kramer et al. (1995) noted that for the KNM-ER 1470/1813 comparison, most distances (62–97%) between analog sample pairs that exceeded the fossil distance were between male-female pairs. For the KNM-ER 1470/OH 24 comparison, more than 50% of all analog pair distances exceeding the fossil distance were between male-female pairs. Thus, if the fossil pairs were of the same species, then sexual dimorphism could account for the variation between them. Kramer et al. (1995, p. 458) concluded, with respect to the magnitude of variation, that "contrary to Lieberman et al. (1988), and in support of Miller (1994), this study has demonstrated that the degree of variation present in the remains attributed to *H. habilis* is not excessive when compared with that observed within some modern, large-bodied hominoid species." Nevertheless, Kramer et al. (1995) provisionally rejected the null hypothesis of intraspecific variation for the *H. habilis* sample on the basis of their variability profiles discussed above.

Average taxonomic distances. In a similar analysis, Grine et al. (1996) used an exact randomization approach to determine if specimens OH 24, KNM-ER 1470, KNM-ER 1813, Stw 53, and SK 847 could all be included within *H. habilis.* They employed analog samples of 44 modern *H. sapiens,* 50 *P. troglodytes,* 50 *G. gorilla,* 3 robust australopithecine crania (OH 5, KNM-ER 406, and WT 17000), and 2 *H. erectus*-grade specimens (ER 3733 and WT 15000). The average taxonomic (Euclidean) distances (d) were calculated between all possible pairs of early *Homo* fossils and these were compared to the distribution of d values between all possible pairs of specimens for each analog species. One analysis was conducted based on 34 linear measurements, and another on a reduced set of 29 measurements to allow the inclusion of specimen SK 847. For each of these data sets, separate calculations were made for variables in raw form and after conversion to shape variables. Additionally, the Euclidean distance matrix for each data set was assessed by minimum spanning trees and ordination resulting from a principal coordinates analysis (see Ordination of Fossils below).

Figure 7A depicts the data, in histogram form, resulting from the first analysis of 34 variables for raw and shape data. The bars represent the percentage of d values in the analog samples that are greater than the d values between KNM-ER 1470 and each of KNM-ER 1813, OH 24, and Stw 53. Bars reaching above the horizontal line indicate that the distance between the respective fossil pair falls within the 95% "population" limits of the analog sample (i.e., the fossil d distances are not statistically significant). This comparison is somewhat analogous to a t test at a 5% level of significance. The raw data show that distances between KNM-ER 1470 and the other three specimens are "significant" relative to the modern human sample (<5%), less than significant relative to the *P. troglodytes* sample (5–10%), and typical relative to the *G. gorilla* sample (about 50%). When size is "removed," the "shape" distances between KNM-ER 1470 and the other three fossil specimens are not significant relative to any modern analog. Especially interesting is the d value for the shape data between KNM-ER 1470 and KNM-ER 1813, which are arguably the "extremes" of the *H. habilis* sample, that is exceeded by large proportions of pair-wise distances among all analog samples. The results are similar for the analyses based on 29 variables (Fig. 7B). One difference is the

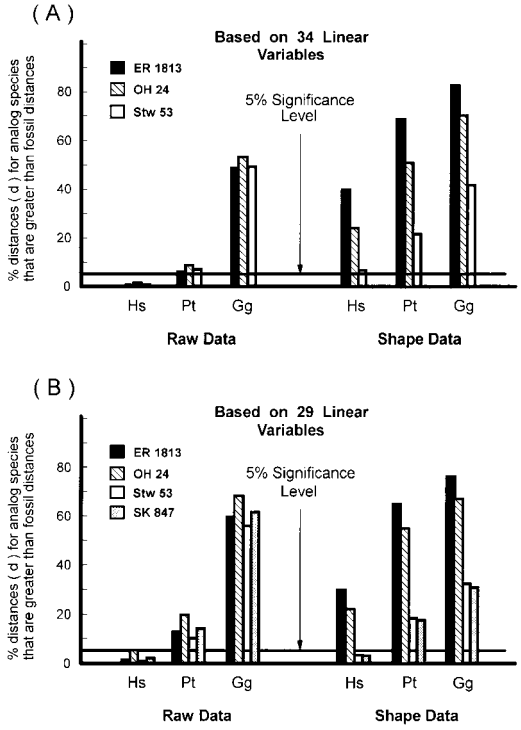

Fig. 7. Exact randomization data of Grine et al. (1996). The vertical bars represent the percentages of average taxonomic distances (*d*) among all possible pairs of modern humans (Hs), *P. troglodytes* (Pt), and *G. gorilla* (Gg) that are greater than the distance between KNM-ER 1470 and other fossil hominids. (**A**) Distances between KNM-ER 1470 and ER 1813, OH 24, and Stw 53 for raw data and shape data based on 34 variables. (**B**) Distances between KNM-ER 1470 and ER 1813, OH 24, Stw 53, and SK 847 for raw data and shape data based on 29 variables.

shape distances between KNM-ER 1470 and Stw 53 and SK 847, which are significantly greater than in the modern human sample. But again, the distances between the fossils are easily matched in the African ape samples.

Given these results, it is curious that Grine et al. (1996, p. 219) concluded that their analysis "suggests that at least two and possibly three taxonomic sets exist within the hypodigm of *H. habilis sensu lato*" (i.e., KNM-ER 1813, KNM-ER 1470, OH 24, SK 847, and Stw 53). In reaching this conclusion, the *G. gorilla* data and its relevancy as an analog were explicitly rejected. Although *P. troglodytes* was not rejected as a relevant analog, their conclusions seem to ignore those data as well.

Grine et al. (1996) argue that it is precisely the comparison with the fossils that demonstrates the irrelevancy of *G. gorilla* as an appropriate analog of intraspecific variability. Their argument is twofold. (1) Gorillas (and other large-bodied hominoids) with large sexually dimorphic canines are not suitable analogs because their facial skeleton is dominated by very large, sexually dimorphic canine roots to a much greater degree than chimpanzees, recent humans, and presumably early hominids. (2) If *G. gorilla* were to be employed as a yardstick of intraspecific variability, not only could crania as different as KNM-ER 1470 and KNM-ER 1813 be accommodated within a single species, but so could specimens as markedly dissimilar as OH 5 and Stw 53 which are generally thought to be different species.

The second argument confuses within-group and between-group variation in the Euclidean data space. This data space, defined by the raw (or shape) variables, is nonisometric in terms of the meaningfulness of distances in different directions relative to within-group versus between-group variation (Albrecht, 1980). Nonisometry affects interpretations of among-group (see Albrecht, 1980) as well as within-group relationships. Consequently, a shorter distance between two specimens could easily represent an interspecific distance, whereas a longer distance could reflect intraspecific variation. Figure 8 shows a simple bivariate example of such a situation. The two clouds represent different species. The interspecific distance between centroids A and B of the two clouds is shorter and has a different probabilistic meaning than the intraspecific distance between points C and D. This can be easily generalized to a multivariate example. Exact randomization methods using distances from nonisometric data spaces do not take this into account since they treat all distances equally. Consequently, interspecific distances, such as that between OH 5 and Stw 53, could easily be obscured in such an analysis which mistakenly treats all distances in a nonisometric data space as though they were all equally meaningful in a probabilistic sense. This is what gives rise to the lament of Grine et al. (1996) regard-

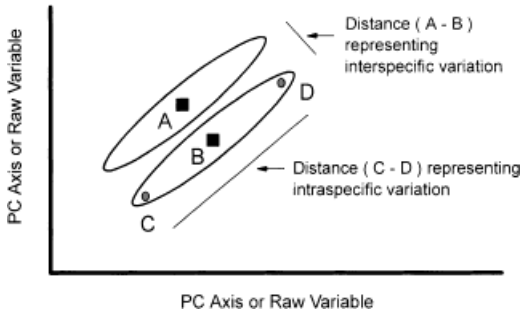

Fig. 8. A bivariate example of the problems of comparing within-group and among-group distances in a nonisometric data space (see Albrecht, 1980). The two ellipses represent the scatter of specimens for two different species in a nonisometric data space (e.g., raw data or principal components). The group centroid of each species is represented by the squares labeled "A" or "B." Two specimens of species "B" are labeled "C" and "D." The interspecific distance between the two species centroids is less than the intraspecific distance separating specimens "C" and "D." The interpretation of these distances is dependent on the patterns of within-group relative to between-group variation. The biological and probabilistic meaningfulness of distances in nonisometric data spaces is not solely in their magnitude or length, but also in their orientation or direction. Methods that compare all distances from a nonisometric data space as though they were equally meaningful are thus problematic (as is the case when exact randomization approaches using principal component or average taxonomic distances are used to make probabilistic statements). Techniques like canonical variates analysis or generalized distances that standardize the data space so that all distances are isometric and equally meaningful in all directions avoid some of these problems.

ing the inability of their method to recognize OH 5 and Stw 53 as belonging to different taxa. However, their problem lies in the misinterpretation of distances in a nonisometric data space and not that *G. gorilla* is an inappropriate analog for early hominids.

The first argument about *G. gorilla* not being an appropriate analog on the basis of canine dimorphism is an untested assertion. Without appropriate substantiation, it provides no basis for discounting any large-bodied, sexually dimorphic hominoid as a relevant analog.

Generalized distances. Donnelly (1996) used generalized distances and randomization procedures to examine the difference in facial shape and size between KNM-ER 1470 and KNM-ER 1813. In his analysis, the distances were separated into "size" and "shape" components. These distances between the fossils were then compared to distributions of size and shape distances between individuals in analog samples of modern *H. sapiens, P. troglodytes, P. pygmaeus,* and *G. gorilla.* The regions of the face and neurocranium were analyzed separately.

Donnelly (1996) reported no statistically significant differences in the distance representing "facial shape" when the two hominid fossils are compared to the analogs. When the distance for "facial size" is compared to the analogs, it is only significantly greater ($P < 0.05$) with respect his samples of modern *H. sapiens* and *P. troglodytes,* but is not significantly greater than *G. gorilla* and *P. pygmaeus* samples. In the cranial vault analysis, the shape distance is not significant relative to any analog. However, the cranial size distance is significant relative to the modern *H. sapiens, P. troglodytes,* and *P. pygmaeus* samples, but not when compared to *G. gorilla* ($P < 0.10$). So with respect to size and shape of the face and neurocranium, the distance for the fossil specimens was not significant relative to the *G. gorilla* sample.

Nevertheless, Donnelly (1996), like Grine et al. (1996), apparently discounts the *G. gorilla* data without providing any justification for doing so and concludes that "the results suggest that KNM-ER 1470 and KNM-ER 1813 represent different species" (Donnelly, 1996, p. 99). A detailed consideration of this study must await the full reporting of the results. On the basis of Donnelly's (1996) preliminary report, however, the null hypothesis of intraspecific variation for the H. habilis sample cannot be ruled out.

Ordinations of fossils excluding within-group variation

A very problematic aspect of some studies is the tendency to present data for fossil specimens in graph form without including a proper standard of within-group variation. Figure 9A is an adaptation of a log ratio diagram presented by Rightmire (1993: Fig. 7). This diagram represents the plotted differences between the logged original measurements for ten variables. KNM-ER 1470 is used as the standard and is represented by the straight line at zero. The differences between KNM-ER 1470 and KNM-ER 1813,

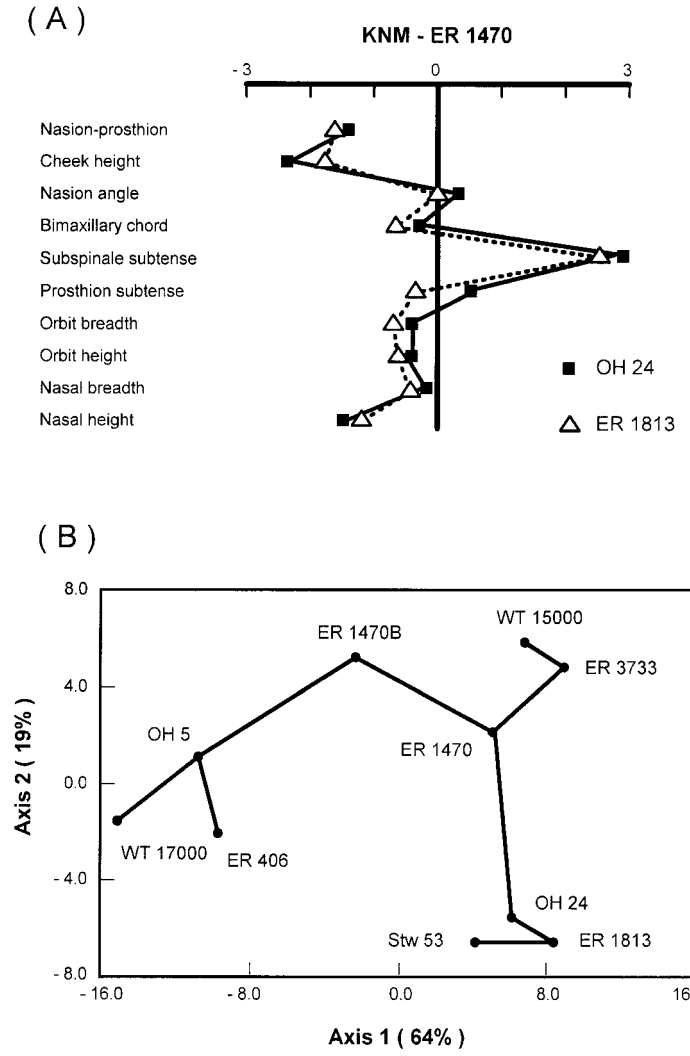

Fig. 9. Two examples of ordinations of fossils that exclude standards of within-group variation. (**A**) Log ratio diagram adapted from Rightmire (1993). OH 24 and KNM-ER 1813 are compared to KNM-ER 1470. The diagram is constructed by plotting the differences in the logged measurements between KNM-ER 1470 and the two other specimens for the variables indicated on the left. (**B**) A principal coordinate diagram adapted from Grine et al. (1996) based on average taxonomic distances. The various fossil specimens are joined together by a minimum spanning tree.

and KNM-ER 1470 and OH 24 are plotted as jagged lines. Although differences, as well as similarities, are noticeable on the diagram, there is no standard of intraspecific variation against which to interpret the meaning of these differences. Without such a standard, as could be obtained through the use of extant referent samples, the data are uninterpretable with respect to the level and meaning of variation that is being sampled. Consequently, one cannot assume that differences between individual fossil specimens indicated by such plots represent anything more than intraspecific variation.

A similar problem is encountered in the study of Grine et al. (1996). Figure 9B is an adaptation of their minimum spanning tree graph from a principal coordinates analysis of an 11 × 11 matrix of average taxonomic distances for some early *Homo* and robust australopithecine fossils (Grine et al., 1996: Fig. 3). They used this graph, and similar ones (e.g., Grine et al., 1996: Figs. 5 and 7), in conjunction with their exact randomization data, and some UPGMA clustering dendrograms (Grine et al., 1996: Figs. 4, 6, and 8) as evidence for the rejection of the single species hypothesis for *H. habilis.* Again, in these ordinations, there is no adequate standard of within-group (intraspecific) variation. Consequently, the distances and relationships among the individual fossils

are uninterpretable. At best, there are a couple of fossils that are considered to represent the same taxon (e.g., OH 5 and ER 406 for *A. boisei,* WT 15000 and ER 3733 for African *H. erectus*). But, these minuscule samples (N = 2/taxon) hardly represent an adequate characterization of intraspecific variation and cannot reasonably be used as any kind of "yardstick" to interpret the distances among other fossils (contra Grine et al., 1996).

In the interpretation of Grine et al.'s (1996) ordinations, there is also the problem of nonisometry in the Euclidean distances depicted in their graphs. There is no way to meaningfully interpret the distances among the fossil specimens without: (1) standardizing the data space (e.g., as is done in canonical variates analysis), or (2) knowing the size, shape, orientation, and relationships of the clouds of intraspecific variation which surround specimens such as OH 5, ER 406, and WT 17000. Such information about the nature of intraspecific and interspecific variation could, of course, be inferred through the appropriate use of extant analog samples. However, this was not done.

Interpretations of fossil ordinations without standards of within-group variation are simply typological in nature. The benefit of quantitative methods, and multivariate methods in particular, is that they allow large amounts of data to be summarized easily at all of the various hierarchical levels of variation (see Albrecht and Miller, 1993: Fig. 1). As such, they should facilitate population thinking, the appropriate use of extant analogs, and the interpretation of variation among fossil specimens in a biologically meaningful manner.

DISCUSSION

The present study is significant in that it is the first time that the breadth of evidence used for recognizing multiple species in the *H. habilis* cranial sample has been independently evaluated. The main result of this critical review and reexamination is that, contrary to previous thought, there is no firm basis on which to reject the null hypothesis of intraspecific variation. Interestingly, while recognition of multiple species in the *H. habilis* sample seems to increase, the data do not support such recognition. Indeed, the data, at this point, are actually quite consistent in showing that the *H. habilis* sample is neither too great in degree, nor too different in pattern of variation to warrant rejection of the single species hypothesis.

Does this mean that it has been proved here that the *H. habilis* sample represents a single species? No. Failure to reject a null hypothesis does not constitute proof of that hypothesis. Indeed, hypotheses can never be "proven" only supported, modified, or rejected. However, at present, the null hypothesis of intraspecific variation does explain the data satisfactorily by means of the present tests. This does not preclude further testing and the possibility of rejecting the null hypothesis in the future.

The implications of this study are several. Firstly, this study demonstrates the value of independent critical reexamination of data and methods. Science is a self-correcting process precisely because independent researchers critically evaluate the data, methods, and conclusions of their peers. In doing so, advances are made in the appropriate use of analytical techniques and data interpretation. Credit must be given to those researchers who, though perhaps making some methodological and/or interpretative missteps, nevertheless attempt to bring a new sophistication to the application of quantitative techniques to this problem of fossil species recognition (e.g., probability estimates of sexual dimorphism, CV profiles, CVA, exact randomization, among others). Yet, it is the independent, critical reevaluation of these approaches that allows the actual consistency of the data regarding the variation in the *H. habilis* sample to be revealed and clears the way for further methodological advances and refinements in future tests of the null hypothesis.

Secondly, a significant difficulty in most of the studies reviewed above is the small size of the *H. habilis* cranial sample. It is, perhaps, the problem of small sample size that has driven researchers to increasingly sophisticated analytical methods in order to "squeeze" more information out of the few available fossil specimens. But increasing

methodological sophistication, however laudable and necessary, cannot substitute for a much-needed, larger, fossil sample. Increased field research targeting Plio-Pleistocene deposits is desperately needed to increase the fossil sample of early *Homo* specimens. It is only with a much larger fossil sample that the nature of morphological variation among specimens can be more rigorously elucidated.

Thirdly, along with methodological advances, there is a need for increasing sophistication with respect to characterizing morphological variation in modern analog species. These, after all, are the "yardsticks" against which the fossils are compared. As the questions regarding the nature of variation in fossil samples become more refined (e.g., in trying to discern whether a sample represents a single species or a mixture of two closely related taxa), then, of necessity, the analysis of intra- and interspecific variation in modern analogs must become more refined. A step in this direction is to recognize the hierarchical nature of variation (see Albrecht and Miller, 1993) within and between modern species and then to apply this more in-depth understanding to the interpretation of variation in fossil samples (e.g., see Miller et al., 1997, 1998). Unless a "yardstick" is well-calibrated with respect to these hierarchical levels of variation (e.g., intraspecific variation including intersubspecific, intrasubspecific, interdemic, intrademic, intersexual, intrasexual or individual variation), then its utility in making finer-grained distinctions in decisions regarding intra- versus interspecific variation becomes compromised. The study of morphological variation in modern analog species (be they hominoids or non-hominoids) is interesting on its own merits, let alone, of immense value for the purpose of fossil species recognition.

Fourthly, this reappraisal of the data makes clear the great need for caution in the recognition of new taxa. A case in point is the recognition of "*Homo rudolfensis*" (Groves, 1989; Wood, 1991, 1993) whose reality as a species is thrown into serious question, at best, by the results of the present study. This is quite apart from any question regarding whether the correct procedures were or were not followed in its actual naming (see Kennedy, 1999, and Wood, 1999). Yet, "*H. rudolfensis*" has already taken on a life of its own, and its recognition provokes other dependent questions such as the nature of the evolutionary relationships between the "two" early *Homo* taxa and other hominids (e.g., Lieberman et al., 1996; Strait et al., 1997). When sample sizes are small, caution and hesitancy should be the rule in recognizing new taxa until such time as the alpha taxonomy is more securely based on larger samples. This would perhaps avoid the reification of nonexistent species.

Fifthly, the finding of the present study only pertains to the *H. habilis* cranial sample. It says nothing about the situation regarding post-crania and isolated dental remains. These latter two are indeed separate samples that deserve separate treatment by investigators. And, it would not be surprising if heterogeneity were to be detected in one of these two other samples. The post-cranial sample, in particular, is quite suspect given the disparity of such specimens such as OH 62 and KNM-ER 3735 (see Leakey et al., 1989) which appear very australopithecine-like and KNM-ER 1472 and KNM-ER 1481A which appear very *Homo*-like. It would not be surprising if the early *Homo* post-cranial sample were contaminated inadvertently by some australopithecine specimens. Indeed, with respect to OH 62, White's first inclination was to recognize it as a small, female, robust australopithecine (see Johanson and Shreeve, 1989). Perhaps this initial hypothesis deserves revisiting.

The question of heterogeneity should be suspected whenever sympatric species are known. In the time period of from 2.4–1.6 mya, there are indeed several sympatric hominid species known of the genera *Homo* and *Australopithecus* (or *Paranthropus*). Procedurally, this means that if heterogeneity were demonstrated unequivocally in the *H. habilis* sample, then its contamination by a recognized species, such as *A. boisei,* would be the next appropriate area of investigation. Ruling out this possibility is prerequisite to recognizing new species.

CONCLUSIONS

This study has critically examined the broad scope of evidence used to support the claim of multiple species in the *H. habilis* cranial sample. Contrary to previous thought, the present reanalysis demonstrates that the null hypothesis of intraspecific variation cannot yet be reasonably rejected. Thus, the recognition of multiple early *Homo* taxa, at present, is currently unwarranted. This conclusion should not be taken as an endorsement of the belief that all the specimens that have ever been attributed to *H. habilis* actually belong in that taxon. Likewise, this conclusion should not be misinterpreted as if it were a claim that only a single species is present. It is simply a recognition that currently, the data do not allow us to go further. This does not preclude a future rejection of the null hypothesis of intraspecific variation, but rather encourages further testing through the recovery of additional fossils, more rigorous analysis, and a better understanding of the nature of hierarchical variation within and between species.

ACKNOWLEDGMENTS

My sincerest thanks to Gene H. Albrecht (University of Southern California) and Bruce R. Gelvin (California State University, Northridge) for all their guidance and encouragement. I also thank Philip Tobias (University of the Witwatersrand) for some helpful comments, R. W. Thorington and staff of the National Museum of Natural History (Smithsonian), Emöke Szathmáry (AJPA editor), and the Department of Pathology/Laboratory Medicine at UCLA.

APPENDIX A. Mean ECVs and mean Ids for hominoid samples

Taxon	Sample #	Series/reference	Males Mean ECV	Males N	Females Mean ECV	Females N	Mean Id
P. troglodytes	1	Selenka (1899)	420	24	390	26	1.08
	2	Zuckerman (1928)	399.5	34	365.8	27	1.09
	3	Ashton and Spence (1958)	410	33	380	78	1.08
	4	Schultz (1965)	381	56	350	57	1.09
	5	Schultz (1965)	356	6	329	5	1.08
G. gorilla	1	Selenka (1899)	510	50	450	48	1.13
	2	Randall (1943–44)	543	133	461	78	1.18
	3	Ashton and Spence (1958)	550	63	460	50	1.20
	4	Schultz (1965)	535	72	443	43	1.21
	5	Harris (1926)	512.3	23	442.2	11	1.16
P. pygmaeus	1	Selenka (1898, 1899)	455	80	390	70	1.17
	2	Openheim (1911–12)	395	?	357.6	?	1.10
	3	Gaul (1933)	434.1	36	389.8	59	1.11
	4	Ashton and Spence (1958)	415	30	370	18	1.12
	5	Schultz (1965)	416	57	338	52	1.23
	6	Hrdlicka (1907)	432.5	4	351.4	7	1.23
H. sapiens (modern)	1	Congo series; Benington (1911–12)	1343.9	47	1205.9	21	1.11
	2	Gaboon, 1864; Benington (1911–12)	1380.5	49	1231.7	43	1.12
	3	Gaboon, 1880; Benington (1911–12)	1447.4	16	1240.4	17	1.17
	4	Bidford-on-Avon; Brash and Young (1935)	1553	6	1378	5	1.13
	5	Kerma, Egypt; Collet (1933)	1383.3	72	1229.9	47	1.12
	6	Esquimaux; Duckworth (1896)	1509	5	1322.5	2	1.14
	7	Mori-Ori; Duckworth (1900)	1434.3	7	1370	3	1.05
	8	Naqada, Egypt; Collet (1933)	1381	88	1287.9	123	1.07
	9	Andaman Isles; Flower (1880)	1244.5	11	1127.5	12	1.10
	10	Timor-Laut; Garson (1884)	1603.8	4	1311	5	1.22
	11	Maiden Castle; Goodman and Morant (1940)	1514.7	19	1380.7	19	1.1
	12	Farringdon St.; Hooke (1926)	1481.5	86	1296.5	132	1.14
	13	Teita Hills, Kenya; Kitson (1931)	1481.5	30	1192.4	33	1.1
	14	Castle Hill; Little (1943–46)	1513.9	39	1365	15	1.11
	15	Whitechapel; MacDonell (1904)	1476.9	72	1299.9	80	1.14
	16	Moorfields; MacDonell (1906–07)	1473.8	22	1365.3	31	1.08
	17	Nepalese; Morant (1924)	1436.2	47	1249.2	6	1.15
	18	First Dynasty, Egypt; Morant (1925)	1364.3	24	1217.8	6	1.12
	19	Australia A; Morant (1927)	1294.6	146	1147.4	67	1.13

APPENDIX A. (continued)

Taxon	Sample #	Series/reference	Males Mean ECV	Males N	Females Mean ECV	Females N	Mean Id
	20	Tasmanian; Morant (1927)	1264.3	33	1153.8	25	1.10
	21	Spitalfields; Morant and Hoadley (1931)	1415.7	68	1245.7	15	1.14
	22	Short Cist, Scottish; Reid and Morant (1928)	1492.5	18	1411.3	6	1.06
	23	Zulu; Ricklan and Tobias (1986)	1373.3	50	1251.2	50	1.1
	24	Bushmen; Shrubsall (1898)	1309.1	9	1252.9	14	1.04
	25	Badarian, Egypt; Stoessiger (1927)	1370.7	35	1274.1	22	1.08
	26	Hythe, England; Stoessiger and Morant (1932)	1456.3	110	1318	83	1.1
	27	Torres Straits, Oceania; Thomas (1885)	1433.2	18	1258	19	1.14
	28	Burmese A; Tildesday (1921)	1406.9	27	1267.9	27	1.11
	29	Burmese B; Tildesday (1921)	1415	4	1232.4	11	1.15
	30	Burmese C; Tildesday (1921)	1442.2	5	1231.4	7	1.17
	31	Easter Isle, Oceania; von Bonin (1931)	1462.2	36	1304.8	26	1.12
	32	New Britain, Oceania; von Bonin (1936)	1285.7	55	1207.7	30	1.06
	33	Nineth Dynasty, Egypt; Woo (1930–31)	1426.9	35	1252.5	25	1.14
	34	Tasmanian, Oceania; Wunderly (1939)	1247.1	14	1242.8	14	1.00
	35	West Scottish; Young (1931)	1460.5	506	1328.6	364	1.1

APPENDIX B. Cranial angles for H. habilis *and hominoid analogs*[1]

Specimens/analogs	Facial triangle NAA	Facial triangle BAA	Facial triangle PRA	Anterior cranial triangle NBA	Anterior cranial triangle BRA	Anterior cranial triangle BBA	Transverse facial angles SSA	Transverse facial angles NFA
Cranial angles								
OH 24	97 (87)	40 (44)	43 (49)	70.5 (69)	41.5 (56.5)	68 (54)	151 (147)	175 (167)
ER 1813	82 (88)	40 (43)	58 (49)	69 (66)	57 (56)	54 (58)	140.5 (147)	142 (150)
ER 1470	85 (86)	45 (44)	50 (50)	68 (62)	57 (65)	55 (53)	153 (155)	153 (155)
Observed *H. habilis* range	15 (2)	5 (1)	15 (1)	2.5 (7)	15.5 (9.5)	14 (5)	20.5 (14)	33 (17)
Normal range for analogs (95% population limits calculated as +/−2 S.D.s from the mean)								
H. sapiens (modern)	14	9	8	7	4	6	18	10
P. troglodytes	18	9	13	6	8	6	27	18
G. gorilla	26	16	15	12	20	17	16	22
P. pygmaeus	23	15	15	17	14	12	28	20

[1] Stringer (1986) from Figs 4–7, My data are in parentheses.

LITERATURE CITED

Albrecht GH. 1980. Multivariate analysis and the study of form, with special reference to canonical variate analysis. Am Zool 20:679–693.

Albrecht GH. 1992. Assessing the affinities of fossils using canonical variates and generalized distances. Hum Evol 7:49–69.

Albrecht GH, Miller JMA. 1993. Geographic variation in primates: a review with implications for interpreting fossils. In Kimbel WH, Martin LB, editors. Species, species concepts, and primate evolution. New York: Plenum Press. p 123–161.

Albrecht GH, Miller JMA. 1997. Variability profiles in fossil studies: the case of *Homo habilis.* Am J Phys Anthropol Suppl 24:65.

Ashton EH, Spence TF. 1958. Age changes in the cranial capacity and foramen magnum of hominoids. Proc Zool Soc Lond 130:169–181.

Benington RC. 1911–1912. A study of the negro skull with special reference to the Congo and Gaboon crania. Biometrika 8:292–339.

Bilsborough A, Wood BA. 1988. Cranial morphometry of early hominids: facial region. Am J Phys Anthropol 76:61–86.

Brash JC, Layard D, Young M. 1935. The Anglo-Saxon skulls from Bidford-On-Avon, Warwickshire and Burwell, Cambridgeshire, with a comparison of their principal characters and those of the Anglo-Saxon skulls in London museums. Biometrika 27:373–387.

Bromage TG. 1992. The ontogeny of *Pan troglodytes* craniofacial architectural relationships and implications for early hominids. J Hum Evol 23:235–252.

Bromage TG. 1993. Developmental criteria for an empirically based reconstruction of KNM-ER 1470. Am J Phys Anthropol Suppl 16:63.

Chamberlain AT. 1987. A taxonomic review and phylogenetic analysis of *Homo habilis.* Ph.D. Thesis. University of Liverpool.

Collett M. 1933. A study of twelfth and thirteenth dynasty skulls from Kerma (Nubia). Biometrika 25:254–284.

Cope DA. 1993. Measures of dental variation as indicators of multiple taxa in samples of sympatric Cerco-

pithecus species. In: Kimbel WH, Martin LB, editors. Species, species concepts, and primate evolution. New York: Plenum Press. p 211–237.
Deinard A, Kidd K. 1999. Evolution of a HOXB6 intergenic region within the great apes and humans. J Hum Evol 36:687–703.
Deinard AS, Sirugo G, Kidd KK. 1998. Hominoid phylogeny: inferences from A sub-terminal minisatellite analyzed by repeat expansion detection (RED). J Hum Evol 35:313–317.
Donnelly SM. 1996. How different are KNM-ER 1470 and KNM-ER 1813? A multivariate comparison using randomization methods. Am J Phys Anthropol Suppl 22:99.
Duckworth WLH. 1896. Notes on a collection of crania of esquimaux, exhibited by Prof. A. Macalister, F. R. S. J Anthopol Inst 9:141–152.
Duckworth WLH. 1900. On a collection of crania, with two skeletons, of the Mori-Ori, or aborigines of the Chatham Islands. With a note on some crania from the same islands now in the Museum of the Royal College of Surgeons. J Anthropol Inst 9:141–152.
Flower WH. 1880. On the osteology and affinities of the natives of the Andaman Islands. J Anthropol Inst 9:108–135.
Garson JG. 1884. On the cranial characters of the natives of Timor-Laut. J Anthropol Inst 13:386–403.
Gaul G. 1933. Uber die wachstumveranderungen am gehirn-schadel des orang-utan. Z Morphol Anthropol 31:362–394.
Goodman CN, Morant GM. 1939. The human remains of the iron age and other periods from Maiden Castle, Dorset. Biometrika 31:295–312.
Grine FE, Jungers WL, Schultz J. 1996. Phenetic affinities among early *Homo* crania from East and South Africa. J Hum Evol 30:189–225.
Groves CP. 1989. A theory of human and primate evolution. New York: Oxford University Press.
Harris HA. 1926. Endocranial form of gorilla skulls with special reference to the existence of dolichocephaly as a normal feature of certain primates. Am J Phys Anthropol 9:157–172.
Harvey PH, Clutton-Brock TH. 1985. Life history variation in primates. Evolution 39:559–581.
Harvey PH, Martin RD, Clutton-Brock TH. 1987. Life histories in comparative perspective. In Smuts BB, Cheney DL, Seyfarth RM, Wrangham RW, Struhsaker TT, editors. Primate societies. Chicago: University of Chicago Press. p 181–196.
Holloway RL. 1973. New endocranial values for the East African early hominids. Nature (London) 243: 97–99.
Holloway RL. 1978. Problems of brain endocast interpretation and African hominid evolution. In: Jolly CJ, editor. Early hominids of Africa. New York: St. Martin's Press. p 379–401.
Holloway RL. 1983. Human brain evolution: a search for units, models and synthesis. Can J Anthropol 3:215–230.
Hooke BGE. 1926. A third study of the English skull with special reference to the Farringdon Street crania. Biometrika 18:1–55.
Howell FC. 1978. Hominidae. In: Maglio VJ, Cooke HBS, editors. Evolution of african mammals. Cambridge, MA: Harvard University Press. p 154–258.
Howells WW. 1973. Cranial variation in man. Papers of the Peabody Museum, Harvard University. 67:1–259.
Howells WW. 1989. Cranial variation in man. Papers of the Peabody Museum, Harvard University. 79:1–189.
Howells WW. 1995. Cranial variation in man. Papers of the Peabody Museum, Harvard University. 82:1–108.
Hrdlicka A. 1907. Anatomical observations on a collection of orang skulls from western Borneo; with a bibliography. Proc US Nat Mus 31:539–568.
Johanson D, Shreeve J. 1989. Lucy's child, The discovery of a human ancestor. New York: William Morrow.
Johanson DC, Masao FT, Eck GG, White TD, Walter RC, Kimbel WH, Asfaw B, Manega P, Ndessokia P, Suwa G. 1987. New partial skeleton of *Homo habilis* from Olduvai Gorge, Tanzania. Nature (London) 327: 205–209.
Kelley J, Plavcan JM. 1998. A simulation test of hominoid species number at Lufeng, China: Implications for the use of the coefficient of variation in paleotaxonomy. J Hum Evol 35:577–596.
Kennedy GE. 1999. Is "*Homo rudolfensis*" a valid species? J Hum Evol 36:119–121.
Kimbel WH, Martin LB. 1993. Species, species concepts, and primate evolution. New York: Plenum Press.
Kimbel WH, Rak Y. 1993. The importance of species taxa in paleoanthropology and an argument for the phylogenetic concept of the species category. In: Kimbel WH, Martin LB, editors. Species, species concepts, and primate evolution. New York: Plenum Press. p 461–484.
Kitson E. 1931. A study of the negro skull with special reference to the crania from Kenya colony. Biometrika 23:271–314.
Kramer A, Donnelly SM, Kidder JH, Ousley SD, Olah SM. 1995. Craniometric variation in large-bodied hominoids: testing the single-species hypothesis for *Homo habilis.* J Hum Evol 29:443–462.
Leakey MD, Clarke RJ, Leakey LSB. 1971. New hominid skull from Bed I, Olduvai Gorge, Tanzania. Nature (London) 232:308–312.
Leakey REF, Walker A. 1988. New *Australopithecus boisei* specimens from Lake Turkana, Kenya. Am J Phys Anthropol 76:1–24.
Leakey RE, Walker A, Ward CV, Grausz HM. 1989. A partial skeleton of a gracile hominid from the upper Burgi member of the Koobi Fora formation, East Lake Turkana, Kenya. In: Hominidae: Proceedings of the 2nd International Congress on Human Paleontology. Milan: Editoriale Jaca Book. p 167–173.
Lieberman DE, Pilbeam DR, Wood BA. 1988. A probabilistic approach to the problem of sexual dimorphism in *Homo habilis:* A comparison of KNM-ER 1470 and KNM-ER 1813. J Hum Evol 17:503–511.
Lieberman DE, Wood BA, Pilbeam DR. 1996. Homoplasy and early *Homo:* an analysis of the evolutionary relationships of *H. habilis* sensu stricto and *H. rudolfensis.* J Hum Evol 30:97–120.
Little KL. 1943–1946. A study of a series of human skulls from Castle Hill, Scarborough. Biometrika 33: 25–35.
Lockwood CA, Richmond BG, Jungers WL, Kimbel WH. 1996. Randomization procedures and sexual dimorphism in *Australopithecus afarensis.* J Hum Evol 31: 537–548.
MacDonell WR. 1904. A study of the variation and correlation of the human skull, with special reference to English crania. Biometrika 2:191–244.
MacDonell WR. 1906–1907. A second study of the English skull, with special reference to Moorfields crania. Biometrika 5:86–104.
Marks J. 1993. Hominoid heterochromatin: terminal C-bands as a complex, genetic trait linking chimpanzee and gorilla. Am J Phys Anthropol 90:237–246.
Marks J. 1994. Blood will tell (won't it?): a century of molecular discourse in anthropological systematics. Am J Phys Anthropol 94:59–79.
McHenry HM. 1991. Sexual dimorphism in *Australopithecus afarensis.* J Hum Evol 20:21–32.

Miller JA. 1990. *Homo habilis:* the sexual dimorphism hypothesis reconsidered. Am J Phys Anthropol 81: 269.
Miller JA. 1991. Does brain size variability provide evidence of multiple species in *Homo habilis?* Am J Phys Anthropol 84:385–398.
Miller JMA. 1994. *Homo habilis:* is the degree of craniofacial variation excessive? Am J Phys Anthropol Suppl 18:147.
Miller JMA. 1995. Craniofacial variation in *Homo habilis:* a multivariate comparison of KNM-ER 1470 and KNM-ER 1813. Am J Phys Anthropol Suppl 20:154.
Miller JMA, Gelvin BR. 1993. *Homo habilis* and small robust australopithecine mandibles: a case of confusion? Am J Phys Anthropol Suppl 16:146.
Miller JMA, Albrecht GH, Gelvin BR. 1997. A hierarchical analysis of craniofacial variation in *Homo habilis* using a gorilla analog. Am J Phys Anthropol Suppl 24:170.
Miller JMA, Albrecht GH, Gelvin BR. 1998. A hierarchical analysis of craniofacial variation in *Homo habilis* using a modern human analog. Am J Phys Anthropol Suppl 26:163.
Morant GM. 1924. A study of certain oriental series of crania including the Nepalese and Tibetan series in the British Museum (Natural History). Biometrika 16:1–105.
Morant GM. 1925. A study of Egyptian craniology from prehistoric to Roman times. Biometrika 17:1–52.
Morant GM. 1927. A study of the Australian and Tasmanian skulls, based on previously published measurements. Biometrika 19:417–440.
Morant GM, Hoadley MF. 1931. A study of the recently excavated Spitalfields crania. Biometrika 23:191–248.
Oppenheim S. 1911–1912. Zur typologie des primatencraniums. Z Morph Anthropol 14:1–204.
Oxnard C. 1973. Form and pattern in human evolution. Chicago: The University of Chicago Press.
Oxnard CE. 1975. Uniqueness and diversity in human evolution. Morphometric studies of Australopithecines. Chicago: The University of Chicago Press.
Oxnard CE. 1983a. Sexual dimorphisms in the overall proportions of primates. Am J Primatol 4:1–22.
Oxnard CE. 1983b. The order of man. New Haven, CT: Yale University Press.
Oxnard CE. 1987. Fossil, teeth, and sex. Seattle: University of Washington Press.
Randall FE. 1943–1944. The skeletal and dental development and variability of the gorilla. Hum Biol 15: 236–254, 307–337; 16:23–76.
Reader J. 1981. Missing links, The hunt for earliest man. Boston: Little, Brown, and Company.
Reid RW, Morant GM. 1928. A study of the Scottish Short Cist crania. Biometrika 20:379–388.
Ricklan DE, Tobias PV. 1986. Unusually low sexual dimorphism of endocranial capacity in a Zulu cranial series. Am J Phys Anthopol 71:285–293.
Richmond BG, Jungers WL. 1995. Size variation and sexual dimorphism in *Australopithecus afarensis* and living hominoids. J Hum Evol 29:229–245.
Rightmire GP. 1993. Variation among early *Homo* from Olduvai Gorge and the Koobi Fora region. Am J Phys Anthropol 90:1–33.
Rogers J. 1993. The phylogenetic relationships among *Homo, Pan,* and *Gorilla:* A population genetics perspective. J Hum Evol 25:201–215.
Schultz AH. 1965. The cranial capacity and the orbital volume of hominoids according to age and sex. Homenaje a Juan Comas en su 65 anniversario, Vol. II. Editorial Libros de Mexico.
Selenka E. 1898. Studien uber entwickelungsgeschichte der tiere. Heft 6: Menschenaffen (Anthropomorphae). Studien uber entwickelung und schadelbau. I. Rassen, Schadel und Bezahnung des Orangutan. Wiesbaden: C. W. Kreidel. p 1–91.
Selenka E. 1899. Studien uber entwickelungsgeschichte der tiere. Heft 7: Menschenaffen (Anthropomorphae). Studien uber entwickelung und schadelbau. II. Schadel des gorilla und schimpanse. Wiesbaden: C. W. Kreidel. p 95–160.
Shrubsall F. 1898. Crania of African bush races. J Anthropol Inst 27:263–292.
Sokal RR, Braumann CA. 1980. Significance tests for coefficients of variation and variability profiles. Syst Zool 29:50–66.
Stoessiger BN. 1927. A study of the Badarian crania recently excavated by the British School of Archaeology in Egypt. Biometrika 19:110–150.
Stoessiger BN, Morant GM. 1932. A study of the crania in the vaulted ambulatory of Saint Leonard's Church, Hythe. Biometrika 24:135–202.
Strait DS, Grine FE, Moniz MA. 1997. A reappraisal of early hominid phylogeny. J Hum Evol 32:17–82.
Stringer CB. 1986. The credibility of *Homo habilis.* In: Wood B, Martin L, Andrews P, editors. Major topics in primate and human evolution. Cambridge: Cambridge University Press. p 266–294.
Suwa G, White TD, Howell FC. 1996. Mandibular postcanine dentition from the Shungura Formation, Ethiopia: Crown morphology, taxonomic allocations, and Plio-Pleistocene hominid evolution. Am J Phys Anthropol 101:247–282.
Thomas O. 1885. Account of a collection of human skulls from Torres Straits. J Anthropol Inst 14:328–343.
Tildesley ML. 1921. A first study of the Burmese skull. Biometrika 13:176–262.
Tobias PV. 1971. The brain in hominid evolution. New York: Columbia University Press.
Tobias PV. 1975. Brain evolution in the hominoidea. In: Tuttle RH, editor. Primate functional morphology and evolution. Paris: Mouton Publishers. p 353–392.
Tobias PV. 1987. The brain of *Homo habilis:* A new level of organization in cerebral evolution. J Hum Evol 16:741–761.
Tobias PV. 1991. Olduvai Gorge, Vol 4: The skulls, endocasts and teeth of *Homo habilis.* Cambridge: Cambridge University Press.
Uchida A. 1992. Intra-species variation and sexual dimorphism in *Gorilla gorilla.* Am J Phys Anthropol Suppl 14:164–165.
von Bonin G. 1931. A contribution to the craniology of the Easter Islanders. Biometrika 23:249–270.
von Bonin G. 1936. On the craniology of Oceania. Crania from New Britain. Biometrika 28:123–148.
Walker A, Leakey REF. 1978. The hominids of East Turkana. Sci Am 239:54–66.
Walker AC, Leakey RE. 1988. The evolution of *Australopithecus boisei.* In: Grine FE, editor. Evolutionary history of the "robust" Australopithecines. New York: Aldine de Gruyter. p 247–258.
White TD, Johanson DC, Kimbel WH. 1983. *Australopithecus africanus* its phyletic position reconsidered. In: Ciochon RL, Corruccini RS, editors. New interpretations of ape and human ancestry. New York: Plenum Press. p 721–780.
Woo TL. 1930–1931. A study of seventy-one ninth dynasty Egyptian skulls from Sedment. Biometrika 22: 65–93.
Wood BA. 1975. An analysis of sexual dimorphism in primates. Ph.D. thesis, University of London.
Wood BA. 1978. Classification and phylogeny of East African hominids. In: Chivers DJ, Joysey KA, editors.

Recent advances in primatology. New York: Academic Press. p 351–372.
Wood B. 1985. Early *Homo* in Kenya, and its systematic relationships. In: Delson E, editor. Ancestors: The hard evidence. New York: Alan R. Liss. p 206–214.
Wood B. 1991. Koobi Fora Research Project, Vol. 4. Oxford: Clarendon Press. p 230–254.
Wood B. 1992a. Early hominid species and speciation. J Hum Evol 22:351–365.
Wood B. 1992b. Origin and evolution of the genus *Homo.* Nature (London) 355:783–790.
Wood B. 1993. Early *Homo*: how many species? In: Kimbel WH, Martin LB, editors. Species, species concepts, and primate evolution. New York: Plenum Press. p 485–522.
Wood B. 1999. "*Homo rudolfensis*" Alexeev, 1986—fact or phantom? J Hum Evol 36:115–118.
Wood BA, Yu Li, Willoughby C. 1991. Intraspecific variation and sexual dimorphism in cranial and dental variables among higher primates and their bearing on the hominid fossil record. J Anat 174:185–205.
Wood BA, Wood C, Konigsberg L. 1994. *Paranthropus boisei:* An example of evolutionary stasis? Am J Phys Anthropol. 95:117–136.
Wunderly J. 1939. The cranial and other skeletal remains of Tasmanians in collections in the Commonwealth of Australia. Biometrika 30:305–337.
Yaroch LA, Thompson JL. 1996. The thin-plate spline and paleoanthropology: an innovative technique for measuring biological shape. Am J Phys Anthropol Suppl 22:249.
Young M. 1931. The West Scottish skull and its affinities. Biometrika 23:10–22.
Zar JH. 1984. Biostatistical analysis. Englewood Cliffs, NJ: Prentice-Hall.
Zuckerman S. 1928. Age-changes in the chimpanzee, with special reference to growth of brain, eruption of teeth, and estimation of age; with a note on the Taungs ape. Proc Zool Soc Lond Part 1:1–42.